Cost-effectiveness of psychotherapy for Borderline Personality Disorder

Cost-Effectiveness of Psychotherapy for Borderline Personality Disorder:

Advances in Assessment and Analysis

Pim Wetzelaer

ISBN: 978-90-829118-5-5

Cost-Effectiveness of Psychotherapy for Borderline Personality Disorder:

Advances in Assessment and Analysis

Pim Wetzelaer

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Maastricht, op gezag van de Rector Magnificus, Prof. dr. Rianne M. Letschert volgens het besluit van het College van Decanen, in het openbaar te verdedigen op donderdag 13 december 2018 om 16.00 uur.

Promotores:

Prof. dr. A. Arntz

Prof. dr. mr. S.M.A.A. Evers

Prof. dr. M.M. Rijkeboer

Beoordelingscommissie:

Prof. dr. M.A. Joore (voorzitter)

Prof. dr. J.J. van Busschbach (Erasmus MC Rotterdam)

Prof. dr. D.P. Bernstein

Dr. J. Bosmans (Vrije Universiteit, Amsterdam)

Dr. J. Lobbestael

The studies presented in this thesis were supported by ZonMw and Mind Netherlands:

CONTENTS

Chapter 1

General introduction

1.1 Introduction

This thesis presents a collection of studies that were performed to advance the assessment and analysis of the cost-effectiveness of psychotherapy for borderline personality disorder. This chapter serves as a general background to the studies. The nature and consequences of BPD are described, as well as the treatments that are currently considered as optimal for BPD. Also, the relevance of performing research on the economic aspects of interventions for BPD is addressed, followed by an explanation of the basic scientific terminology and underlying methods that are used to formally interpret the term 'cost-effectiveness'. Finally a brief introduction is given of each of the studies that are presented in the subsequent chapters of this thesis.

1.2 Borderline personality disorder

1.2.1 A large burden

Borderline personality disorder (BPD) is a mental health disorder with potentially devastating consequences for an individual's personal life. It may affect patients' emotional as well as physical well-being, their daily functioning in general, their plans, ambitions, professional careers, as well as their ability to interact with others. The diagnostic label BPD is applied when at least five of the following nine symptom criteria are met, in a way that is characteristic for a person's general functioning since early adulthood (or earlier): 1) fear of abandonment, 2) unstable interpersonal relationships, 3) uncertain self-image or identity, 4) impulsive behaviour, 5) self-injurious behaviour, 6) emotional changeability or hyperactivity, 7) feelings of emptiness, 8) difficulty controlling intense anger, 9) transient suspiciousness or dissociation (*DSM-IV-TR: Diagnostic and statistical manual of mental disorders, text revision*, 2000).

The prevalence of BPD in the Netherlands has been estimated at 1.1% in the general population (ten Have et al., 2016). In other countries, similar BPD prevalence estimates range between 0.5 and 2.7% (Samuels, 2011). Several studies have indicated that the quality of life of patients with BPD is severely impaired (Feenstra et al., 2012; IsHak et al., 2013; Perseius et al., 2006; Soeteman, Verheul, & van Busschbach, 2008). A diagnosis of BPD is a large burden to bear for a patient, and often also for his or her family, friends, colleagues, care providers, and others in his or her environment.

To society as a whole, the burden of BPD is significant in economic terms as well. To a large extent this can be explained by an extensive use of health care services, including both inpatient and outpatient facilities (Bender et al., 2001; Coid et al., 2009; Feenstra et

al., 2012; Soeteman, Hakkaart-van Roijen, et al., 2008). Adding these costs to the costs of productivity losses (van Asselt et al., 2007; Soeteman, Hakkaart-van Roijen, et al., 2008) as well as other costs, such as those related to informal care and out-of-pocket costs (van Asselt et al., 2007), the total societal costs for BPD are substantial.

1.2.2 Treatment of BPD

There are four types of specialized psychotherapy that have been found to be effective for BPD (Zanarini, 2009): dialectical behaviour therapy (DBT; Linehan, 1993), schema therapy (ST; Arntz & Van Genderen, 2011; Young et al., 2003), mentalization based treatment (MBT; Bateman & Fonagy, 2004), and transference-focused psychotherapy (TFP; Clarkin & Kernberg, 2015). These are intensive, long-term treatments that are delivered by specially trained psychotherapists, either in individual sessions only (e.g., TFP), or as a combination of individual and group psychotherapy (e.g., DBT and MBT). The high number of sessions that are needed, in particular those consisting of individual psychotherapy, can make these interventions costly.

These four types of specialized psychotherapy are currently the recommended treatments for BPD in the Netherlands (Netwerk kwaliteitsontwikkeling GGz, 2018). However, only between 2% and 51% (median = 23%) of patients who are diagnosed with BPD at intake in Dutch specialized mental health centres receive psychotherapy as a first-line treatment (Hermens et al., 2011). It is important to note that psychotherapy in this context does not even refer to specialized psychotherapy for BPD. This suggests that many patients with BPD currently receive suboptimal care (e.g., only supportive care or pharmacological treatment to provide symptom relief) and that there is substantial room for extension of the supply of specialized psychotherapy for BPD in the Netherlands.

One specific type of specialized psychotherapy for BPD is schema therapy, which aims at full recovery (Arntz & Van Genderen, 2011; Young et al., 2003). Schemas refer to the set of cognitive representations that originate in early childhood and are the basis for beliefs about the self, others, and the world. Moreover, they are associated with the expression of certain behaviours and emotions. The so-called 'early maladaptive schemas' develop as the result of unmet needs during childhood (e.g., core emotional needs such as trust, love, attention, and security). Although adaptive at first, these schemas can become maladaptive later in life when they lead to unhealthy behaviour. Schema therapy makes use of a variety of therapeutic techniques targeted at reducing the expression of maladaptive schemas and associated dysfunctional coping patterns. Recently, a group psychotherapy format has been developed for schema therapy (Farrell et al., 2012). A group setting offers a supportive environment of peers that allows patients to learn from each other and practice their healthy behaviour. Furthermore, it provides therapists with

an additional therapeutic angle that is not present in an individual setting. For example, the group setting may offer a family-like context in which 'limited re-parenting' (i.e., a quintessential ingredient of schema therapy is that the therapist acts as a good parent for the patient, within professional limits) can take place. Such factors could increase the effectiveness of the intervention. Previous research on group schema therapy (GST) consists of a study that investigated 8 months of GST as an addition to treatment as usual (Farrell et al., 2009), and a pilot study that investigated a combination of individual and group schema therapy (Dickhaut & Arntz, 2014). The results of both studies are promising. However, neither has compared GST as a standalone treatment for BPD to treatment as usual. In addition, the cost-effectiveness of GST has not yet been assessed. If schema therapy can indeed be effectively provided in a group format, then that would suggest a substantial improvement in the efficiency of health care resource use in comparison to individual psychotherapy.

1.3 Economic evaluation

1.3.1 A worrisome future

In 2011 the total costs of mental health care in the Netherlands amounted to € 5,700,000,000. In the same year a report was issued by the 'CPB Netherlands Bureau for Economic Policy Analysis' that gave insight into how the costs of Dutch health care have developed over time in the forty years prior to its publication (van der Horst et al., 2011). With no exception, each year they were higher than the year before. Relative to the Dutch national income (i.e., the Gross Domestic Product (GDP)), the total of health care costs has increased from 8% in 1972 to 13% in 2010. This rise can largely be explained by health care getting more expensive itself as well as demographic factors such as an increased life expectancy. Moreover, it was alarmingly noted that if the health care costs in the Netherlands will continue to rise at this pace, the costs would amount to an estimated 19 - 31% of the GDP in 2040. Somewhat reassuringly, the rise has been less strong in 2013, 2014, and 2015 in comparison to the fifteen years before (Centraal Bureau voor de Statistiek, 2016). Relative to the Dutch GDP, the health care costs have decreased for three consecutive years to a level of 14% in 2015.

Relative to other health care sectors, the trend of rising costs is the most prominent for mental health care (Bijenhof et al., 2012). Between 1998 and 2010, mental health care costs have more than doubled. For the treatment of personality disorders specifically, the increase in total health care costs has been even stronger, in yet a shorter period of time.

1.3.2 Turning the tide

Given the prospect of health care costs outstretching national health care budgets, the making of choices is imminently necessary. In order to keep health care systems sustainable for the future, health care interventions need to be compared based on their value for money or, in other words, their cost-effectiveness (Drummond et al., 2015). Such comparisons may then inform health insurers or policy makers, who decide whether an intervention is reimbursed or not. This will reduce the provision of health care interventions that cannot be demonstrated to improve health outcomes in a way that justifies their costs. As such, it is an evidence-based approach to an efficient allocation of health care budgets.

The successful treatment of BPD implies a substantial alleviation from its burden to both patients and society. For the individual patient, an effective psychotherapy reduces the severity of symptoms to non-pathological levels and restores impairments in quality of life. For society, the benefits of investing in such treatments include a reduction in the use of other health care services, as well as reductions in other societal costs (e.g., productivity losses, informal care and out-of-pocket costs). The question that remains is whether the investment in psychotherapy, either in terms of the extension of the supply of existing interventions or in terms of the adding of new interventions to current treatment options, can be considered worthwhile. The studies that are presented in this thesis share the common aim of contributing to an answer to this important question.

1.3.3 A brief explanation of terms and methods

In a *cost-effectiveness analysis* (CEA) two or more interventions can be compared based on their (potential) differences in costs (i.e., the *incremental costs* or ΔC) and differences in effectiveness (i.e., *incremental effects* or ΔE). This is what is typically expressed as an *incremental cost-effectiveness ratio* (ICER):

$$ICER = \frac{C_2 - C_1}{E_2 - E_1} = \frac{\Delta C}{\Delta E}.$$

For an intervention to be considered as more cost-effective than another, it is important that any differences in effectiveness reasonably weigh up against any differences in costs. More formally this means that the ICER should not exceed a certain limit, or so-called ceiling ratio. This limit is defined as the willingness-to pay (WTP) for one additional unit of effectiveness. To avoid the statistical difficulties in the interpretation of ratios, a linear reformulation of the ICER has been proposed (Hoch et al., 2002), referred to as the incremental net monetary benefit (INMB):

$$INMB = \lambda * (E_2 - E_1) - (C_2 - C_1) = \lambda * \Delta E - \Delta C,$$

where λ is the willingness-to-pay value. This furthermore has the advantage that net benefits (NBs) can be calculated using individual patient level data following $NB_i = \lambda * E_i - C_i$, which can then be used as the dependent variable in statistical regression (e.g., see Chapter 6).

The outcome measures that are used in scientific research to assess the clinical effectiveness of an intervention can either be disorder-specific (e.g., based on symptom severity) or generic (i.e., meaning that the same outcome measure can be used for a broad range of disorders, including both physical as well as mental disorders). A cost-effectiveness analysis can be performed based on both types of outcomes measures (i.e., in addition to costs). A generic outcome measure that is often used to quantify a patient's quality of life is the *utility* value. These are expressed as values ranging between 1 (full health) and 0 (dead). When changes in quality of life are repeatedly assessed over time, this makes the calculation of quality-adjusted life years (QALYs) possible. In case the latter are used in an analysis to assess cost-effectiveness, it is (technically) called a *cost-utility analysis* (CUA). Both a CEA as well as a CUA are examples of an *economic evaluation* (Evers et al., 1997).

For the interpretation of the results from a CUA, or in other words, to assess the cost-effectiveness of an intervention based on its 'costs per QALY gained', a ceiling ratio is required. The ceiling ratio that is applied to a specific disorder, depends on the *burden of disease*. Similar (yet inverse to) utility values, these are expressed as values ranging between 0 (no burden) and 1 (largest possible burden). In the Dutch guideline (Zwaap et al., 2015) the following ceiling ratios are provided: € 20,000 per QALY for a burden of disease of 0.1 - 0.4, € 50,000 per QALY for a burden of disease of 0.41 - 0.7, and € 80,000 per QALY for a burden of disease of 0.71 - 1.0. For BPD, the estimated burden of disease is 0.54 (Vos & Mathers, 2000). Therefore, a ceiling ratio of € 50,000 per QALY is used to assess the cost-effectiveness of interventions for BPD.

Although it can be interesting to compare the cost-effectiveness of two specific interventions head-to-head, often it makes more sense to assess whether a new, or say, experimental intervention has any additional value in comparison to the optimal, state-of-the-art interventions that patients usually receive (e.g., as recommended by clinical guidelines). The latter are often referred to as *treatment as usual* (TAU).

In scientific investigations designed to assess the clinical effectiveness and cost-effectiveness of interventions, patients are randomly assigned to either an experimental or control (e.g., TAU) intervention. This is done in order to enable the causal attribution of any differences in relevant outcomes over time to the specific treatments that patients received. Such studies are referred to as *randomized controlled trials* (RCTs).

In order to ensure a sufficient number of participants in an RCT (i.e., to provide adequate statistical power for analysis), often they are performed using multiple treatment centres. Such studies are called *multicentre* RCTs. Also, RCTs are often performed over a longer period of time during which patients are repeatedly assessed. Such studies are referred to as *longitudinal* RCTs. Both multicentre and longitudinal RCTs introduce a hierarchical, or 'nested' structure to the data. In multicentre RCTs, participants are nested in treatment centres. In longitudinal RCTs, the repeated assessments are nested in participants. An approach that is specifically designed for the statistical analysis of nested data is *multilevel modelling*. Multilevel modelling also provides an efficient approach to the handling of missing values in longitudinal data. Other often used approaches to the handling of cases with missing values include complete case analysis (i.e., cases with missing values are excluded from analysis) or imputation-based approaches.

An important reason for the use of complete case analysis or imputation-based methods for the handling of cases with missing values, is because the total costs (i.e., the sum of costs that are assessed at different points in time) of individual patients are needed to perform a bootstrap analysis. This is a method to approximate the sampling distribution of a variable by means of a resampling procedure. Often this is done non-parametrically, which means that no specific statistical distribution (e.g., a normal distribution) is specified beforehand. The advantage of such an approach is that it facilitates the accommodation of *skewed* data, or in other words, a variable with an asymmetric probability density function. In health economics, right-skewed cost data are common due to the fact that it is usually only a small number of patients who incur relatively high costs (e.g., due to crises or adverse events), whereas a higher number of patients incur relatively low costs. In addition, costs are by definition always positive. An alternative, parametric approach to dealing with skewed cost data is by assuming gamma or lognormal data distributions for costs (e.g., see Chapters 3, 5 and 6 of this thesis for examples).

Finally, there are two different approaches to statistical analysis: the frequentist approach and the Bayesian approach. The latter is more attractive in the context of an economic evaluation since it allows the estimation of a 'probability of cost-effectiveness' instead of calculating the probability of finding the observed sample data under the assumption that an intervention is not cost-effective (i.e., by performing a null hypothesis test). The frequentist approach defines a probability as a limiting long-run frequency. This is suitable for an application to events that are repeatable, such as rolling dice or tossing coins to test their fairness. The Bayesian approach defines a probability as a degree of belief. This can be applied to things or events about which one is uncertain. Some authors argue that the Bayesian definition fits better with the interpretation of a parameter in a statistical model that expresses the relative cost-effectiveness of two particular interventions (i.e., such as in a net benefit regression), since it is generally specific

to the problem being studied (O'Hagan & Luce, 2003). In other words, they interpret such a parameter as a one-off or non-repeatable event that is not subject to any random variability. Instead, the uncertainty regarding its value is caused by not knowing it.

1.4 This thesis

In short, this thesis consists of a systematic literature review, a model-based economic evaluation, (the study protocol and preliminary results of) a trial-based economic evaluation, and a methodological study.

First, a systematic review is presented in Chapter 2 that provides an up-to-date overview of the economic evaluation literature regarding psychotherapy for personality disorders (Wetzelaer et al., 2016). In particular, the general characteristics of the included studies and the specific characteristics of the economic evaluations, including an assessment of their quality, are presented and summarized.

In Chapter 3, a model-based economic evaluation is presented with the aim to demonstrate a method for the synthesis of empirical evidence on the costs, effectiveness, and cost-effectiveness of specialized psychotherapy for borderline personality disorder (Wetzelaer et al., 2017). Specifically, the empirical evidence that is obtained after performing a systematic literature review is used as an input for the simulation of patient-level data. These are then synthesized to assess the cost-effectiveness and budget impact of the further extension of the supply of specialized psychotherapy for borderline personality disorder in the Netherlands.

The study protocol for an international, multicentre RCT on group schema therapy (GST) for borderline personality disorder (Wetzelaer et al., 2014) is presented in Chapter 4, which includes an economic evaluation. In this study two different formats of GST, one that consists of only group psychotherapy and one that consists of a combination of group and individual psychotherapy, and TAU are compared. The details regarding the design of the study, including patient recruitment, the scheduling of assessments, the interventions, and the outcome measures and instruments that are used for the assessment of clinical effectiveness and costs, are presented and discussed.

In Chapter 5, the preliminary results of the economic evaluation described in Chapter 4 are reported, based upon the available data from the treatment centres in the Netherlands. Additionally, a detailed description of the methods used for the assessment and analysis of the data on costs and effectiveness, the CEA, and the CUA are provided. Due to their preliminary nature, the results presented in this chapter are blinded as to the interventions to which they pertain. For the same reason, the work presented in this chapter is not amenable for publication.

In Chapter 6, a method for the analysis of longitudinal cost-effectiveness data using Bayesian multilevel models as an extension of the net benefit regression framework is presented. The idea behind this approach is that it is an efficient way of making use of a dataset that includes cases with missing values, without having to exclude those cases or the need for additional procedures for the imputation of missing data. A coherent set of models for the development of net benefit over time is presented, as well as how it can be determined which model fits the data best. To demonstrate the method, it is applied to an empirical example.The results from the best fitting model are presented, as well as those of variants that assume lognormal and gamma distributed data. Furthermore, a series of appendices to this manuscript is provided, including a description of the statistical details of the method, as well as an extensive manual that presents the model code and the commands to run the models.

Chapter 7 concludes with a general discussion. It consists of a critical reflection upon the work that was performed for this thesis as well as directions for the possible future work to build further upon the insights gathered over the course of my Ph.D. trajectory.

1.5 References

Arntz, A., & Van Genderen, H. (2011). *Schema therapy for borderline personality disorder*. John Wiley & Sons.

Bateman, A., & Fonagy, P. (2004). Mentalization-based treatment of BPD. *Journal of personality disorders*, *18*(1), 36.

Bender, D. S., Dolan, R. T., Skodol, A. E., Sanislow, C. A., Dyck, I. R., McGlashan, T. H., … Gunderson, J. G. (2001). Treatment utilization by patients with personality disorders. *American Journal of psychiatry*, *158*(2), 295-302.

Bijenhof, A. M., Folkertsma, M. A., Kommer, G. J., Slobbe, L. C. J., & Polder, J. J. (2012). *KVZ 2012-1 Kosten Ontwikkeling GGZ.* Rijksinstituut voor Volksgezondheid en Milieu.

Centraal Bureau voor de Statistiek. (2016). *www.cbs.nl/nl-nl/nieuws/2016/20/zorguitgaven-stijgen-langzamer.*

Clarkin, J. F., & Kernberg, O. F. (2015). *Transference-focused psychotherapy for borderline personality disorder: A clinical guide*. American Psychiatric Pub.

Coid, J., Yang, M., Bebbington, P., Moran, P., Brugha, T., Jenkins, R., … Ullrich, S. (2009). Borderline personality disorder: health service use and social functioning among a national household population. *Psychol Med*, *39*(10), 1721-31.

Dickhaut, V., & Arntz, A. (2014). Combined group and individual schema therapy for borderline personality disorder: A pilot study. *J Behav Ther Exp Psychiatry*, *45*(2), 242-251.

Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes*. Oxford university press.

DSM-IV-TR: Diagnostic and statistical manual of mental disorders, text revision. (2000). Washington, DC: American Psychiatric Association.

Evers, S. M., Van Wijk, A. S., & Ament, A. J. (1997). Economic evaluation of mental health care interventions. a review. *Health Econ*, *6*(2), 161-77.

Farrell, J., Shaw, I., & Arntz, A. (2012). *Group schema therapy for borderline personality disorder: A step-by-step treatment manual with patient workbook*. Wiley.

Farrell, J., Shaw, I. A., & Webber, M. A. (2009). A schema-focused approach to group psychotherapy for outpatients with borderline personality disorder: A randomized controlled trial. *Journal of Behavior Therapy and Experimental Psychiatry*, *40*(2), 317-328.

Feenstra, D. J., Hutsebaut, J., Laurenssen, E. M., Verheul, R., Busschbach, J. J., & Soeteman, D. I. (2012). The burden of disease among adolescents with personality pathology: quality of life and costs. *J Pers Disord*, *26*(4), 593-604.

Hermens, M. L., van Splunteren, P. T., van den Bosch, A., & Verheul, R. (2011). Barriers to implementing the clinical guideline on borderline personality disorder in the netherlands. *Psychiatric services*, *62*(11), 1381–1383.

Hoch, J. S., Briggs, A. H., & Willan, A. R. (2002). Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Econ*, *11*(5), 415-30.

IsHak, W. W., Elbau, I., Ismail, A., Delaloye, S., Ha, K., Bolotaulo, N. I., … Wang, C. (2013). Quality of life in borderline personality disorder. *Harv Rev Psychiatry*, *21*(3), 138-50.

Linehan, M. (1993). *Cognitive-behavioral treatment of borderline personality disorder*. Guilford Press.

Netwerk kwaliteitsontwikkeling GGz. (2018). *www.ggzstandaarden.nl.*

O'Hagan, A., & Luce, B. R. (2003). *A primer on Bayesian statistics in health economics and outcomes research.*

Perseius, K. I., Andersson, E., Asberg, M., & Samuelsson, M. (2006). Health-related quality of life in women patients with borderline personality disorder. *Scand J Caring Sci*, *20*(3), 302-7.

Samuels, J. (2011). Personality disorders: epidemiology and public health issues. *Int Rev Psychiatry*, *23*(3), 223-33.

Soeteman, D. I., Hakkaart-van Roijen, L., Verheul, R., & van Busschbach, J. J. (2008). The economic burden of personality disorders in mental health care. *Journal of Clinical Psychiatry*, *69*(2), 259-265.

Soeteman, D. I., Verheul, R., & van Busschbach, J. J. (2008). The burden of disease in personality disorders: diagnosis-specific quality of life. *J Pers Disord*, *22*(3), 259-68.

ten Have, M., Verheul, R., Kaasenbrood, A., van Dorsselaer, S., Tuithof, M., Kleinjan, M., & de Graaf, R. (2016). Prevalence rates of borderline personality disorder symptoms: a study based on the Netherlands mental health survey and incidence study-2. *BMC psychiatry*, *16*(1), 249.

van Asselt, A. D., Dirksen, C. D., Arntz, A., & Severens, J. L. (2007). The cost of borderline personality disorder: societal cost of illness in BPD-patients. *Eur Psychiatry*, *22*(6), 354-61.

van der Horst, A., van Erp, F., & de Jong, J. (2011). *CPB Policy Brief 2011/11.* Centraal Planbureau.

Vos, T., & Mathers, C. D. (2000). The burden of mental disorders: a comparison of methods between the Australian burden of disease studies and the Global Burden of Disease study. *Bulletin of the World Health Organization*, *78*(4), 427-438.

Wetzelaer, P., Farrell, J., Evers, S., Jacob, G. A., Lee, C. W., Brand, O., … Arntz, A. (2014). Design of an international multicentre RCT on group schema therapy for borderline personality disorder. *BMC Psychiatry*, *14*(1), 319.

Wetzelaer, P., Lokkerbol, J., Arntz, A., van Asselt, A., & Evers, S. (2016). [Cost-effectiveness of psychotherapy for personality disorders; a systematic review on economic evaluation studies]. *Tijdschr Psychiatr*, *58*(10), 717-727.

Wetzelaer, P., Lokkerbol, J., Arntz, A., van Asselt, T., Smit, F., & Evers, S. (2017). Cost-effectiveness and budget impact of specialized psychotherapy for borderline personality disorder: A synthesis of the evidence. *Journal of Mental Health Policy and Economics*, *20*(4), 177–190.

Young, J. E., Klosko, J. S., & Weishaar, M. E. (2003). *Schema therapy: A practitioner's guide*. Guilford Press.

Zanarini, M. C. (2009). Psychotherapy of borderline personality disorder. *Acta Psychiatrica Scandinavica*, *120*(5), 373-377.

Zwaap, J., Knies, S., Van der Meijden, C., Staal, P., & Van der Heiden, L. (2015). *Kosten-effectiviteit in de praktijk.* Zorginstituut Nederland.

Chapter 2

Cost-effectiveness of psychotherapy for personality disorders; a systematic literature review of economic evaluation studies

Pim Wetzelaer[1], Joran Lokkerbol[2,3], Arnoud Arntz[1,4], Thea van Asselt[5,6] & Silvia Evers[2,7]

[1]Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands
[2]Centre of Economic Evaluation, Trimbos Institute (Netherlands Institute of Mental Health and Addiction), Utrecht, The Netherlands
[3]Rob Giel Research Center, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
[4]Department of Clinical Psychology, University of Amsterdam, Amsterdam, The Netherlands
[5]Department of Pharmacy, University of Groningen, Groningen, The Netherlands
[6]Department of Epidemiology, University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands
[7]Department of Health Services Research, CAPHRI School of Public Health and Primary Care, Faculty of Health Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands

Based on: Wetzelaer, P., Lokkerbol, J., Arntz, A., van Asselt, A., & Evers, S. (2016). Kosteneffectiviteit van psychotherapie voor persoonlijkheidsstoornissen; een systematisch literatuuronderzoek van economische evaluatiestudies. *Tijdschrift voor Psychiatrie*, 58(10), 717-727.

2.1 Abstract

BACKGROUND Previous economic evaluation studies have been performed that contribute to the evidence base for the cost-effectiveness of psychotherapy for personality disorders. An overview of these studies is lacking.

AIM To provide an overview of the scientific literature on the cost-effectiveness of psychotherapy for patients with a personality disorder.

METHOD A systematic review of the literature was performed, searching the NHS EED, PubMed and PsycINFO databases. Only full economic evaluations of treatments in which all patients had a personality disorder were taken into account.

RESULTS Most studies concluded that at least one of the psychotherapeutic treatments investigated was cost-effective. Dialectical behavior therapy was studied the most; schema therapy came next, followed by cognitive behavioural therapy.

CONCLUSION In general, scientific evidence indicates that psychotherapeutic treatments for patients with personality disorders are cost-effective relative to the comparator treatments. This is important information because it can influence decisions on whether the costs of psychotherapy should be reimbursed.

KEY WORDS cost-effectiveness, personality disorders, psychotherapy

2.2 Introduction

Scientific research increasingly supports the clinical effectiveness of various psychotherapeutic treatments for patients with personality disorders (Bartak et al., 2007). Parallel to clinical studies, economic evaluations shed light on the societal costs and the cost-effectiveness of psychotherapy for personality disorders.

Since psychotherapy for personality disorders is intensive and often requires long-term treatment, the costs are potentially high. However, also the potential savings in costs when a personality disorder is successfully treated can be high, both for the health care sector, as well as other sectors such as the labour market. In the current Dutch health care system, psychotherapeutic treatment of personality disorders is not always (fully) reimbursed. Psychotherapy can therefore be difficult to obtain for patients, it is only provided to a limited extent, or it is denied.

The question therefore rises if, and to what extent, investment in psychotherapeutic treatment for patients with personality disorders can be considered worthwhile, or in other words (Markowitz, 2015): is psychotherapy for personality disorders cost-effective?

In this paper we report a systematic literature review on economic evaluations of psychotherapy for personality disorders. The aim of our study is to give an overview of the scientific literature on the cost-effectiveness of psychotherapeutic treatments for patients with a personality disorder.

2.3 Methods

2.3.1 Search strategy

For our literature search, we used the NHS Economic Evaluation Database (EED), PubMed and PsycINFO, with additional reference checking of the articles that were retrieved. Previous research has shown that the combined use of NHS EED and PubMed is an appropriate way to identify relevant economic evaluations (Alton et al., 2006). Given the specific research area covered in our study, we considered PsycINFO as an important addition. The NHS EED database of the Centre for Reviews and Dissemination consists of economic evaluations within the area of health and social care and has been updated until December 2014 by weekly searches within the following databases: MEDLINE, EMBASE, CINAHL, PsycINFO and PubMed.

We limited our literature study to economic evaluations of psychotherapeutic treatments for personality disorders exclusively, in which the costs and effects of two or more interventions (including psychotherapy) were compared, and which were published in an international, peer-reviewed journal. No date restrictions were used. Economic evaluations of treatments for patients in forensic settings were excluded because we assumed that the costs of those patients are not (fully) representative for the costs of personality disorders in general. We also excluded editorials, letters and earlier reviews, although the references of the latter were checked to identify additional studies.

The following search terms were used for the NHS EED: 'personality disorder OR personality disorders', and for PubMed and PsycINFO: '(personality disorder OR personality disorders) AND (costs and cost analysis[Mesh Terms] OR economics[Mesh Terms] OR costs and cost analysis[Mesh Terms] OR cost-effectiveness OR economic evaluation)'. Our search retrieved 27 studies from the NHS EED, of which 11 were included, 477 in PubMed, of which 14 were included, and 90 in PsycINFO, of which 9 were included. After removing duplicates and adding the economic evaluations from (Brazier et al., 2006), our search resulted in 18 studies that were included.

2.3.2 Quality criteria and score

We assessed the quality of the economic evaluations based on the following six criteria: synthesis of costs and effects (are costs related to effects?), intention-to-treat (ITT) analysis (are all patients analysed in the condition to which they were randomised?), discounting (are all future costs and benefits discounted?), sensitivity analysis (are the results of a sensitivity analysis quantitatively presented?), primary cost data (are the costs based on primary data collection?) and perspective (is the economic evaluation performed from a societal perspective?). For every question answered with 'yes', 1 point was added to the total quality score.

2.3.3 Presentation of the results

We present the most important characteristics of the included studies in two tables that display the general characteristics (Table 1) and the specific characteristics and quality scores of the economic evaluations (Table 2), respectively. The first author screened all studies for the characteristics shown in Table 1 and 2. The other authors screened one quarter of all studies each, so that all studies were assessed twice.

2.4 Results

We included 18 studies in total. The general characteristics of these studies are shown in Table 1. The specific characteristics and quality scores of the economic evaluations are shown in Table 2.

Table 2.1: General characteristics of the included studies

Reference (Country)	n (PD diagnosis)[1]	Format[2]	Time horizon[3]	Intervention (Setting)	Control (Setting)	Outcome measure[4]
Abbass et al., 2008 (Can, US)	27 (Primary diagnosis not reported; BPD: 44%, OPD: 37%, APD: 33%, PD-NOS: 22%, PPD: 19%, DPD: 7%, NPD: 7%, ASPD: 4%, HPD: 4%)	RCT	2.1 years on average	Intensive short-term dynamic psychotherapy (outpatient)	TAU[5]: minimal-contact supportive psychiatric follow-up (during waiting time until start psychodynamic therapy), no psychotherapy (outpatient)	General psychopathological symptoms (BSI), interpersonal problems (IIP)
van Asselt et al., 2008 (NL)	86 (100% BPD)	RCT	4 years	Schema therapy (outpatient)	Transference-focused psychotherapy (outpatient)[6]	Proportion of recovered patients (BPDSI), QALY
Bamelis et al., 2015 (NL)	320 (APD: 51%, DPD: 11%, OPD: 28%, PPD: 4%, HPD: 1%, NPD: 5%)[7]	RCT	3 years	Schema therapy (outpatient) & Clarification-oriented psychotherapy (outpatient)	TAU: optimal care according to the Dutch multidisciplinary guidelines; Insight-oriented psychotherapy: 42%, Cognitive behaviour therapy: 19%, EMDR[8]: 1.5% (diverse)[7]	Proportion of recovered patients (SCID-II), QALY
Beecham et al., 2006 (UK)	108[9] (Different PD types)[10]	Non-RCT	2 - 2.5 years	Step-down (First inpatient psychoanalytically oriented psychotherapy and group psychotherapy, then outpatient treatment incl. group analytic psychotherapy) & One-stage (inpatient psychoanalytically oriented psychotherapy and group psychotherapy)	General psychiatric care, e.g. supportive therapy (diverse)	General functioning (GAS), general psychopathological symptoms (SCL-90)

Continued on next page

Table 2.1 – continued from previous page

Reference (Country)	n (PD diagnosis)[1]	Format[2]	Time horizon[3]	Intervention (Setting)	Control (Setting)	Outcome measure[4]
Brazier et al., 2006; based on Bateman & Fonagy, 1999 (UK)[11]	44 (100% BPD)	RCT	1 year	Psychoanalytic treatment = MBT[12] (day treatment)	TAU: general psychiatric care, no psychotherapy (diverse)	Parasuicidal events (SSHI), QALY (based on BDI)
Brazier et al., 2006; based on van den Bosch et al., 2002 (NL)[11]	58 (100% BPD of whom 53% with substance use)	RCT[13]	1 year	Dialectical behaviour therapy (outpatient)	TAU: general psychiatric care, no psychotherapy (outpatient)	Parasuicidal events (TBR), QALY (based on BDI)
Brazier et al., 2006; based on Koons et al., 2001 (US)[11]	22 (100% BPD)	RCT[13]	1 year	Dialectical behaviour therapy (outpatient)	TAU: individual cognitive behaviour therapy, psychodynamic or eclectic therapy (outpatient)	Parasuicidal events, QALY (based on BDI)
Brazier et al., 2006; based on Linehan et al., 1991 (US)[11]	44 (100% BPD and chronically suicidal)	RCT	1 year	Dialectical behaviour therapy (outpatient)	TAU: psychotherapy of choice (outpatient)	Parasuicidal events
Brazier et al., 2006; based on Turner, 2000 (US)[11]	24 (100% BPD)	RCT[13]	1 year	Dialectical behaviour therapy-oriented therapy (outpatient)	Client-centered psychotherapy (outpatient)[9]	Parasuicidal events (PHI)
Davidson et al., 2010 (UK)	106 (100% BPD)	RCT	6 years[14]	Cognitive behaviour therapy (outpatient) + TAU (diverse)	TAU: general medical and mental health care, no pscyhotherapy (diverse)	Proportion of recovered patients (SCID-II), suicide attempts & parasuicidal events (ADSHI), QALY

Continued on next page

Table 2.1 – continued from previous page

Reference (Country)	n (PD diagnosis)[1]	Format[2]	Time horizon[3]	Intervention (Setting)	Control (Setting)	Outcome measure[4]
Horn et al., 2015 (NL)	(Primary diagnosis not reported; cluster A PD: 2%, cluster B PD: 11%, cluster C PD: 44%, PD-NOS: 46%)	Non-RCT	3 years	Short-term psychotherapy based on transactional analysis (inpatient)	Other psychotherapies (22% cognitive behaviour therapy, 30% psychodynamic, 42% integrative, 5% not specified; diverse)[5]	QALY
Kvarstein et al., 2013 (Nor)	107 (Primary diagnosis not reported; BPD: 48%, APD: 41%, SPD: 1%, PPD: 15%, NPD: 2%, OPD: 9%, DPD: 7%, PD-NOS: 21%)	RCT	3 years	Step-down (First day treatment with psychodynamic and cognitive group therapy, then outpatient psychotherapy)	Individual psychotherapy (by therapist's preference; mostly psychodynamic or psychoanalytic; outpatient)[9]	General functioning (GAF)
Palmer et al., 2006 (UK)	106 (100% BPD)	RCT	2 years	Cognitive behaviour therapy (outpatient) + TAU (diverse)	TAU: general medical and mental health care, no pscyhotherapy (diverse)	QALY
Pasieczny & Connor, 2011 (Aus)	90 (100% BPD)	Non-RCT	6 months	Dialectical behaviour therapy (outpatient)	TAU: clinical case management (during waiting time until start DBT; outpatient or at home)	Suicide attempts and parasuicidal events
Priebe et al., 2012 (UK)	80 (primary diagnosis not reported; BPD: 99%, APD: 66,3%, DPD: 23,8%, OPD: 47,5%, PPD: 51,3%; all with self-injury)	RCT	1 year	Dialectical behaviour therapy (outpatient)	TAU: usual care as provided by the NHS, 13 of 40 received psychotherapy (diverse)	Parasuicidal events

Continued on next page

Reference (Country)	n (PD diagnosis)[1]	Format[2]	Time horizon[3]	Intervention (Setting)	Control (Setting)	Outcome measure[4]
Ranger et al., 2009 (UK)	52 (58% schizophrenia, 19% schizoaffective disorder, 10% bipolar disorder and 10% BPD, 94% with comorbid PD and 6% with personality problems)	RCT	1 year	Nidotherapy[16] + TAU (assertive outreach treatment) (diverse)	TAU: assertive outreach treatment (specialist care for complex cases) (diverse)	General psychopathological symptoms (BPRS), social functioning (SFQ)
Soeteman et al., 2010 (NL)	Model based on data of 241 patients with a cluster B-PD	Non-RCT[17]	5 years	Psychotherapy (inpatient) & psychotherapy (day treatment)	Psychotherapy (outpatient)[9]	Number of years recovered per patient, QALY
Soeteman et al., 2011 (NL)	Model based on data of 466 patients with a cluster B-PD	Non-RCT[17]	5 years	Psychotherapy (inpatient) & psychotherapy (day treatment)	Psychotherapy (outpatient)[9]	Number of years recovered per patient, QALY

[1] Abbreviations: PD = personality disorder, PPD = Paranoid PD, SPD = schizoid PF, STPD = schizotypical PD, ASPD = antisocial PD, BPD = borderline personality disorder, HPD = Histrionic PD, NPD = narcissistic PD, DPD = dependent PD, APD = avoidant PD, OPD = obsessive-compulsive PD, PD-NOS = PD not otherwise specified

[2] Abbreviations: RCT = randomized controlled trial, Non-RCT = non-randomized controlled trial

[3] As used for the economic evaluation

[4] When no synthesis was performed, only the primary outcome measures are listed, and QALY if applicable. Abbreviations: BSI = Brief Symptom Inventory, IIP = Inventory of Interpersonal Problems, BPDSI = Borderline Personality Disorder Severity Index, QALY = Quality-Adjusted Life Year, SCID-II = Structured Clinical Interview for DSM-IV Axis II Personality Disorders, GAS = Global Assessment Scale, SCL-90 = Symptom Checklist-90-R, TBR = Target Behavior Rating, BDI = Beck Depression Inventory, SSHI = Suicide and Self-Harm Inventory, PHI = Parasuicide History Interview, ADSHI = Acts of Deliberate Self-Harm Inventory, GAF = Global Assessment of Functioning, BPRS = Brief Psychiatric Rating Scale, SFQ = Social Functioning Questionnaire

[5] TAU = treatment as usual

[6] Two or more specialist treatments were compared in this study, thus the treatment listed in this column is in fact not a control condition

[7] Information reported in clinical article (Bamelis et al., 2014)

[8] EMDR = eye movement desensitization and reprocessing

[9] Subsample of patients with complete data

[10] Diagnosis not retrievable (personal communication, J. Beecham)

[11] These economic evaluations were performed by / reported in (Brazier et al., 2006)

[12] MBT = mentalization-based therapy

[13] Costs were estimated using a regression model based on data from other studies

[14] 5 year follow-up only, the economic evaluation of the time period during which treatment was provided is described in (Palmer et al., 2006)

[15] Naturalistic study with a correction using propensity scores

[16] Nidotherapy aims to change a patient's personal environment instead of their symptoms or behaviour (see (Tyrer, 2002)

[17] Non-RCT in combintaion with a Markov model (naturalistic study with a correction using propensity scores)

Table 2.2: Specific characteristics and quality scores of the economic evaluations

Reference	Quality score (0-6)	Synthesis	ITT-analysis[1]	Discounting[2]	Sensitivity analysis[3]	Primary cost data	Perspective[4]	Health care costs[5]	Patient and family costs[5]	Other costs[5]	Cost-effectiveness[6]
Abbass et al., 2008	2	No	Yes	No	No	Yes	Not reported	1 (estimate) and 2 (medication costs only)	No	7	ISTDP is cost-effective
van Asselt et al., 2008	6	Yes	Yes	Yes	Yes	Yes	Societal	1 and 2	3, 4 and 5	7 and 9	ST is cost-effective
Bamelis et al., 2015	6	Yes	Yes	Yes	Yes	Yes	Societal	1 and 2	3, 4 and 5	7, 8 and 9	ST is cost-effective, COP is not
Beecham et al., 2006	2	Yes	No	No	No	Yes	Not reported	1 and 2	6	7, 8 and 9	DBT is cost-effective
Brazier et al., 2006; based on Bateman & Fonagy, 1999	5	Yes	No	n.a.	Yes	Yes[7]	Societal	1 and 2	6	7, 8 and 9	Inconclusive
Brazier et al., 2006; based on van den Bosch et al., 2002	5	Yes	Yes	n.a.	Yes	No	Societal	1 and 2	6	7, 8 and 9	DBT is cost-effective
Brazier et al., 2006; based on Koons et al., 2001	4	Yes	No	n.a.	Yes	No	Societal	1 and 2	6	7, 8 and 9	DBT is not cost-effective

Continued on next page

Table 2.2 – continued from previous page

Reference	Quality score (0-6)	Synthesis	ITT-analysis[1]	Discounting[2]	Sensitivity analysis[3]	Primary cost data	Perspective[4]	Health care costs[5]	Patient and family costs[5]	Other costs[5]	Cost-effectiveness[6]
Brazier et al., 2006; based on Linehan et al., 1991	5	Yes	No	n.a.	Yes	Yes[8]	Societal	1 and 2	6	7, 8 and 9	DBT is cost-effective
Brazier et al., 2006; based on Turner, 2000	4	Yes	No	n.a.	Yes	No	Societal	1 and 2	6	7, 8 and 9	DBT-oriented therapy is cost-effective
Davidson et al., 2010	1	No	No	No	No	Yes	Patient, health and social care, and other care	1 and 2	6	8 and 9	CBT + TAU less costly on average, no significant difference
Horn et al., 2015	6	Yes	Yes	Yes	Yes	Yes	Societal	1 and 2	No	7	Clinical STP based on TA is the most cost-effective
Kvarstein et al., 2013	2	Yes	No	No	No	Yes	Health care	1 and 2	No	9	Inconclusive
Palmer et al., 2006	4	Yes	Yes	Yes	No	Yes	Patient, health and social care, and other care	1 and 2	6	8 and 9	CBT + TAU is not cost-effective
Pasieczny & Connor, 2011	2	No	No	n.a.	No	Yes	Not reported	1 and 2 (A&E only)	No	No	DBT is cost-effective

Continued on next page

Table 2.2 – continued from previous page

Reference	Quality score (0-6)	Synthesis	ITT-analysis[1]	Discounting[2]	Sensitivity analysis[3]	Primary cost data	Perspective[4]	Health care costs[5]	Patient and family costs[5]	Other costs[5]	Cost-effectiveness[6]
Priebe et al., 2012	5	Yes	Yes	n.a.	Yes	Yes	Health and social care	1 and 2	No	7 and 9	Inconclusive
Ranger et al., 2009	3	Yes	No	n.a.	No	Yes	Health, social, voluntary and criminal justice services	1 and 2	No	8 and 9	Nidotherapy is cost-effective
Soeteman et al., 2010	6	Yes	Yes	Yes	Yes	Yes	Societal	1 and 2	No	7	Outpatient and day treatment PT is the most cost-effective
Soeteman et al., 2011	6	Yes	Yes	Yes	Yes	Yes	Societal	1 and 2	No	7	Inpatient and day treatment (ST) PT is the most cost-effective

[1] ITT = Intention-to-treat; only scored as 'Yes' when all randomized participants were analyzed
[2] Not applicable to studies of 1 year or shorter, these were scored as 1 (same as 'Yes')
[3] Only scored as 'Yes' when the results were reported numerically
[4] Perspective as stated, in case of multiple perspectives the widest is listed
[5] 1 = mental health care costs, 2 = costs for other medical care, 3 = patient out-of-pocket costs, 4 = loss of daily activities, 5 = informal care, 6 = acccomodation costs, 7 = productivity losses, 8 = judicial costs, 9 = costs of social services
[6] Comparisons between two experimental treatments (without TAU) only show which treatment is the most cost-effective, comparisons with TAU show the absolute cost-effectiveness
[7] Costs reported in Bateman & Fonagy, 2003, for estimates of supervision costs additional data from other sources were used
[8] Costs reported in Heard, 2000, for estimates of supervision costs additional data from other sources were used
Abbreviations: ISTDP = intensive short-term dynamic psychotherapy, ST = outpatient schema therapy, DBT = outpatient dialectical behaviour therapy, (ST) PT = (short-term) psychotherapy, TA = transactional analysis.

2.4.1 General characteristics

All studies were performed in Western countries, and six were performed in the Netherlands. The sample size of most studies was limited, in ten studies smaller than n=100. Nine studies focused on the treatment of patients with borderline personality disorder exclusively. Dialectical behaviour therapy was studied the most (six times), next to schema therapy (twice) and cognitive behaviour therapy (twice). In two studies different treatment settings (inpatient, day treatment or outpatient) were compared. Furthermore, two studies described a phased treatment in which psychotherapy was first offered in an inpatient or day treatment setting, and subsequently as outpatient treatment. As a control condition, the majority of studies used treatment as usual (TAU), this being the case for eleven studies. Quality-adjusted life years (QALYs; which are often preferred by policy makers due to their generic nature), but also events such as suicide and self-injury (parasuicidal events), recovery from diagnosis, general functioning, and general psychopathological symptoms were recurrently used as outcome measures.

2.4.2 Characteristics of the economic evaluation

In the vast majority of studies a synthesis of costs and effects was performed, in half of the studies an ITT analysis was performed, and in about half of those studies the future costs and benefits were discounted. The majority of studies included a sensitivity analysis to assess the influence of changes in variables or assumptions on the conclusions. In nearly all studies, cost data were based on primary data collection.

Ten studies were described as having been performed from a societal perspective, yet this was not always reflected in the actual costs that were assessed. For the societal perspective it is required that all relevant costs and benefits are taken into account in the analysis, irrespective of to whom those costs or benefits specifically apply. Despite stating a societal perspective in many studies the assessment of costs was limited to health care costs, complemented with other cost categories such as productivity and judicial costs. Patient and and family costs were not included in many studies. Other studies were performed from (even) narrower perspectives or did not state from which perspective they were performed.

2.4.3 Cost-effectiveness

Most studies provided support for the cost-effectiveness of at least one of the studied psychotherapeutic treatments. Of the six studies on dialectical behaviour therapy, four concluded that it was cost-effective. Of the remaining studies, one was inconclusive and

another concluded that dialectical behaviour therapy is not cost-effective. Also the cost-effectiveness of schema therapy is supported by the available scientific literature (two studies). Regarding cognitive behaviour therapy (added to TAU) one study showed that it is not cost-effective and another (follow-up) study was inconclusive. The studies in which treatment settings were compared showed that for patients with a cluster B personality disorder day treatment and outpatient therapy were the most cost-effective and for patients with a cluster C personality disorder therapies in (short term) day treatment or inpatient settings were most cost-effective. These studies were not RCTs however, which means that the possibility of a selection bias cannot be excluded. The studies in which psychotherapy was first offered as inpatient or day treatment and subsequently as outpatient treatment both were inconclusive regarding the cost-effectiveness of those treatments.

2.5 Discussion

In this review we have provided an overview of economic evaluations of psychotherapeutic treatments for patients with personality disorders. From this, a number of important findings emerge. In general, the scientific literature seems to indicate that psychotherapy is cost-effective. The cost-effectiveness of dialectical behaviour therapy, schema therapy and cognitive behaviour therapy has been studied most often.

Previous research (Brazier et al., 2006; Brettschneider et al., 2014) has indicated that economic evaluations performed in the research area of personality disorders are often of moderate quality. In our literature review we therefore assessed the quality of the included studies. Also when only studies of high quality (i.e., with a quality score of 5 or 6) are taken into account, the cost-effectiveness of psychotherapy is supported by the scientific literature. Our results furthermore offer a starting point for formulating ideas that could help to improve the quality of future studies.

As a point of recommendation, economic evaluations could be performed from a (wider) societal perspective, so that all relevant costs and benefits are taken into account. Most studies were not performed from a societal perspective, which raises the possibility that important savings or costs outside of the health care sector (e.g., productivity losses or judicial costs) have not been taken into account. Furthermore, some studies did not state from which perspective costs and benefits were taken into account and in some other studies the stated perspective did not (fully) correspond to the costs and benefits that were actually taken into account.

To facilitate the comparison of treatments for different disorders based on cost-effectiveness in reimbursement decisions, QALYs are usually preferred as an out-

come measure for the effectiveness of treatments (i.e., cost-utility analysis). Unfortunately, this outcome measure was not used in several studies. As a second point of recommendation, future studies could therefore make use of the QALY as an (additional) outcome measure. This could then facilitate the use of the results of the study in reimbursement decisions.

2.6 Conclusion

From our overview of economic evaluations of psychotherapy for personality disorders, a positive picture emerges. In most cases, the results show that it is worthwhile to invest in psychotherapy. This is an important notion; not just for science, but for health care policy making as well. The cost-effectiveness of psychotherapy for personality disorders should be a convincing argument in considering the reimbursement of the costs of these treatments.

2.7 Acknowledgements

This article was created at the initiative of the Dutch 'Kenniscentrum Persoonlijkheidsstoornissen' (in English: 'Knowledge Centre Personality Disorders'), in response to an invitation for a short contribution to a theme issue of the Dutch journal 'Tijdschrift voor Psychiatrie' (in English: 'Journal of Psychiatry') on the societal costs and benefits of psychiatry.

2.8 References

Abbass, A., Sheldon, A., Gyra, J., & Kalpin, A. (2008). Intensive short-term dynamic psychotherapy for DSM-IV personality disorders: a randomized controlled trial. *J Nerv Ment Dis*, *196*(3), 211-6.

Alton, V., Eckerlund, I., & Norlund, A. (2006). Health economic evaluations: how to find them. *Int J Technol Assess Health Care*, *22*(4), 512-7.

Bamelis, L. L., Arntz, A., Wetzelaer, P., Verdoorn, R., & Evers, S. M. (2015). Economic evaluation of schema therapy and clarification-oriented psychotherapy for personality disorders: a multicenter, randomized controlled trial. *J Clin Psychiatry*, *76*(11), e1432-40.

Bamelis, L. L., Evers, S. M., Spinhoven, P., & Arntz, A. (2014). Results of a multicenter randomized controlled trial of the clinical effectiveness of schema therapy for personality disorders. *Am J Psychiatry*, *171*(3), 305-22.

Bartak, A., Soeteman, D. I., Verheul, R., & Busschbach, J. J. (2007). Strengthening the status of psychotherapy for personality disorders: an integrated perspective on effects and costs. *Can J Psychiatry*, *52*(12), 803-10.

Bateman, A., & Fonagy, P. (1999). Effectiveness of partial hospitalization in the treatment of borderline personality disorder: a randomized controlled trial. *Am J Psychiatry*, *156*(10), 1563-9.

Bateman, A., & Fonagy, P. (2003). Health service utilization costs for borderline personality disorder patients treated with psychoanalytically oriented partial hospitalization versus general psychiatric care. *Am J Psychiatry*, *160*(1), 169-71.

Beecham, J., Sleed, M., Knapp, M., Chiesa, M., & Drahorad, C. (2006). The costs and effectiveness of two psychosocial treatment programmes for personality disorder: a controlled study. *Eur Psychiatry*, *21*(2), 102-9.

Brazier, J., Tumur, I., Holmes, M., Ferriter, M., Parry, G., Dent-Brown, K., & Paisley, S. (2006). Psychological therapies including dialectical behaviour therapy for borderline personality disorder: a systematic review and preliminary economic evaluation. *Health Technol Assess*, *10*(35), iii, ix-xii, 1-117.

Brettschneider, C., Riedel-Heller, S., & Konig, H. H. (2014). A systematic review of economic evaluations of treatments for borderline personality disorder. *PLoS One*, *9*(9), e107748.

Davidson, K. M., Tyrer, P., Norrie, J., Palmer, S. J., & Tyrer, H. (2010). Cognitive therapy v. usual treatment for borderline personality disorder: prospective 6-year follow-up. *Br J Psychiatry*, *197*(6), 456-62.

Heard, H. L. (2000). Cost-effectiveness of dialectical behavior therapy in the treatment of borderline personality disorder. *Dissertation of the University of Washington*.

Horn, E. K., Verheul, R., Thunnissen, M., Delimon, J., Goorden, M., Hakkaart-van Roijen, L., … Busschbach, J. J. (2015). Cost-effectiveness of short-term inpatient psychotherapy based on transactional analysis in patients with personality disorder. *J Pers Disord*, 1-19.

Koons, C. R., Robins, C. J., Tweed, J. L., Lynch, T. R., Gonzalez, A. M., Morse, J. Q., … Bastian, L. A. (2001). Efficacy of dialectical behavior therapy in women veterans with borderline personality disorder. *Behavior therapy*, *32*(2), 371-390.

Kvarstein, E. H., Arnevik, E., Halsteinli, V., Ro, F. G., Karterud, S., & Wilberg, T. (2013). Health service costs and clinical gains of psychotherapy for personality disorders: a randomized controlled trial of day-hospital-based step-down treatment versus outpatient treatment at a specialist practice. *BMC Psychiatry*, *13*, 315.

Linehan, M. M., Armstrong, H. E., Suarez, A., Allmon, D., & Heard, H. L. (1991). Cognitive-behavioral treatment of chronically parasuicidal borderline patients. *Arch Gen Psychiatry*, *48*(12), 1060-4.

Markowitz, J. C. (2015). Is psychotherapy for personality disorders worth it? *J Clin Psychiatry*, *76*(11), e1472-3.

Palmer, S., Davidson, K., Tyrer, P., Gumley, A., Tata, P., Norrie, J., … Seivewright, H. (2006). The cost-effectiveness of cognitive behavior therapy for borderline personality disorder: results from the BOSCOT trial. *J Pers Disord*, *20*(5), 466-81.

Pasieczny, N., & Connor, J. (2011). The effectiveness of dialectical behaviour therapy in routine public mental health settings: An Australian controlled trial. *Behav Res Ther*, *49*(1), 4-10.

Priebe, S., Bhatti, N., Barnicot, K., Bremner, S., Gaglia, A., Katsakou, C., … Zinkler, M. (2012). Effectiveness and cost-effectiveness of dialectical behaviour therapy for self-harming patients with personality disorder: a pragmatic randomised controlled trial. *Psychother Psychosom*, *81*(6), 356-65.

Ranger, M., Tyrer, P., Miloseska, K., Fourie, H., Khaleel, I., North, B., & Barrett, B. (2009). Cost-effectiveness of nidotherapy for comorbid personality disorder and severe mental illness: randomized controlled trial. *Epidemiol Psichiatr Soc*, *18*(2), 128-36.

Soeteman, D. I., Verheul, R., Delimon, J., Meerman, A. M., van den Eijnden, E., Rossum, B. V., … Kim, J. J. (2010). Cost-effectiveness of psychotherapy for cluster B personality disorders. *Br J Psychiatry*, *196*(5), 396-403.

Soeteman, D. I., Verheul, R., Meerman, A. M., Ziegler, U., Rossum, B. V., Delimon, J., … Kim, J. J. (2011). Cost-effectiveness of psychotherapy for cluster C personality disorders: a decision-analytic model in the Netherlands. *J Clin Psychiatry*, *72*(1), 51-9.

Turner, R. M. (2000). Naturalistic evaluation of dialectical behavior therapy-oriented treatment for borderline personality disorder. *Cognitive and Behavioral Practice*, *7*(4), 413-419.

Tyrer, P. (2002). Nidotherapy: a new approach to the treatment of personality disorder. *Acta Psychiatr Scand*, *105*(6), 469-71;discussion 471-2.

van Asselt, A. D., Dirksen, C. D., Arntz, A., Giesen-Bloo, J. H., van Dyck, R., Spinhoven, P., … Severens, J. L. (2008). Out-patient psychotherapy for borderline personality disorder: cost-effectiveness of schema-focused therapy v. transference-focused psychotherapy. *Br J Psychiatry*, *192*(6), 450-7.

van den Bosch, L. M., Verheul, R., Schippers, G. M., & van den Brink, W. (2002). Dialectical behavior therapy of borderline patients with and without substance use problems. implementation and long-term effects. *Addict Behav*, *27*(6), 911-23.

Chapter 3

Cost-effectiveness and budget impact of specialized psychotherapy for borderline personality disorder: a synthesis of the evidence

Pim Wetzelaer[1], Joran Lokkerbol[2,3], Arnoud Arntz[1,4], Thea van Asselt[5,6], Filip Smit[2,7,8] & Silvia Evers[2,9]

[1]Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands
[2]Centre of Economic Evaluation, Trimbos Institute (Netherlands Institute of Mental Health and Addiction), Utrecht, The Netherlands
[3]Rob Giel Research Center, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
[4]Department of Clinical Psychology, University of Amsterdam, Amsterdam, The Netherlands
[5]Department of Pharmacy, University of Groningen, Groningen, The Netherlands
[6]Department of Epidemiology, University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands
[7]Department of Public Mental Health, Trimbos Institute (Netherlands Institute of Mental Health and Addiction), Utrecht, The Netherlands
[8]Department of Clinical Psychology, VU University and EMGO Institute for Health and Care Research, Amsterdam, The Netherlands; Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands
[9]Department of Health Services Research, CAPHRI School of Public Health and Primary Care, Faculty of Health Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands

Reference: Wetzelaer, P., Lokkerbol, J., Arntz, A., Smit, F., & Evers, S. (2017). Cost-effectiveness and Budget Impact of Specialized Psychotherapy for Borderline Personality Disorder: A Synthesis of the Evidence. *The Journal of Mental Health Policy and Economics*, 20(4), 177-190.

3.1 Abstract

BACKGROUND Specialized outpatient psychotherapy for patients with borderline personality disorder (BPD) is expected to reduce their use of other health care resources. It is currently unknown to what extent the costs of providing these interventions can be expected to be offset by a reduction in other health care costs in the Netherlands. To establish the cost-effectiveness and budget impact of specialized outpatient psychotherapy, the estimated incremental costs are synthesized with the estimated incremental effects. We have developed a method for the synthesis of all relevant evidence on clinical effectiveness as well as health care resource use.

AIM The aim of this article is to present a method for the synthesis of evidence for cost-effectiveness and budget impact analysis with a specific application to specialized outpatient psychotherapy for borderline personality disorder in the Netherlands.

METHOD A systematic search of the English-language literature is performed to retrieve evidence on the clinical effectiveness and the health care resource use following 12 months of specialized outpatient psychotherapy for borderline personality disorder. The available evidence is used as an input for a model-based economic evaluation. Simulated patient-level data are used to provide overall estimates of the incremental costs and incremental effects, which serve to assess the cost-effectiveness and budget impact of specialized outpatient psychotherapy for borderline personality disorder in the Netherlands.

RESULTS The results indicate that specialized outpatient psychotherapy for BPD can be considered cost-effective and that its scaling up to Dutch national level would require an investment of € 2.367 million (95% C.I.: € 1,717,000 - € 3,272,000) per 1,000 additional patients with BPD. Sensitivity analyses demonstrated the robustness of our findings in light of several uncertain components and assumptions in our calculations, but also their sensitivity to the choice of included studies based on the comparator condition and the assumption of high intervention costs.

DISCUSSION We present a method for the synthesis of evidence from different types of studies in a way that respects the uncertainty surrounding those findings. Limitations of the study pertain to the inclusion of findings from studies with suboptimal designs, the transferability of research findings, and uncertainty regarding the time horizon considered. More research is needed on the sensitivity of our findings to the choice of included

studies based on the comparator condition.

IMPLICATIONS FOR HEALTH CARE PROVISION AND USE The results suggest that the provision of specialized outpatient psychotherapy for BPD leads to a reduction in other health care resource use. Overall, the results are promising and encourage future studies on aspects that are currently still uncertain.

IMPLICATIONS FOR HEALTH POLICY The results may support policy makers in deciding whether or not to allocate health care budget for the provision of specialized outpatient psychotherapy for patients with BPD in the Netherlands.

IMPLICATIONS FOR FURTHER RESEARCH The results provide important directions for future research. This includes the need for future studies to make a comparison between specialized outpatient psychotherapy and treatment as usual and to have longer follow-up time.

3.2 Introduction

Borderline personality disorder (BPD) is a severe mental disorder leading to unstable functioning in the interpersonal, emotional, cognitive and behavioural domains (Leichsenring et al., 2011). It has a prevalence of around 1.5% in the general population (Widiger, 2012). Psychological crises, self-harm and suicide attempts are common in BPD and often result in hospital admissions. Patients with BPD are furthermore known to make extensive use of mental health services when seeking treatment (Bender et al., 2001; Soeteman et al., 2008). Therefore, BPD imposes substantial economic costs on national health care budgets. On the one hand, these costs could be alleviated by providing specialized outpatient psychotherapy to patients with BPD; on the other hand, additional costs can be expected for providing such interventions.

For the treatment of BPD, four types of specialized psychotherapy have been found to be effective in reducing BPD psychopathology and symptoms (Zanarini, 2009): dialectical behaviour therapy (DBT; Linehan, 1993), schema therapy (ST; Arntz & Van Genderen, 2011), mentalization-based treatment (MBT; Bateman & Fonagy, 2004) and transference-focused psychotherapy (TFP; Clarkin & Kernberg, 2015). Some studies suggest that the costs of providing specialized outpatient psychotherapy are offset by reductions in the costs of other health care services (e.g., Heard, 2000; Wagner et al., 2014). This would imply an overall reduction in health care costs, so that the intervention can be labeled cost-saving from a health care provider's perspective. Since specialized treat-

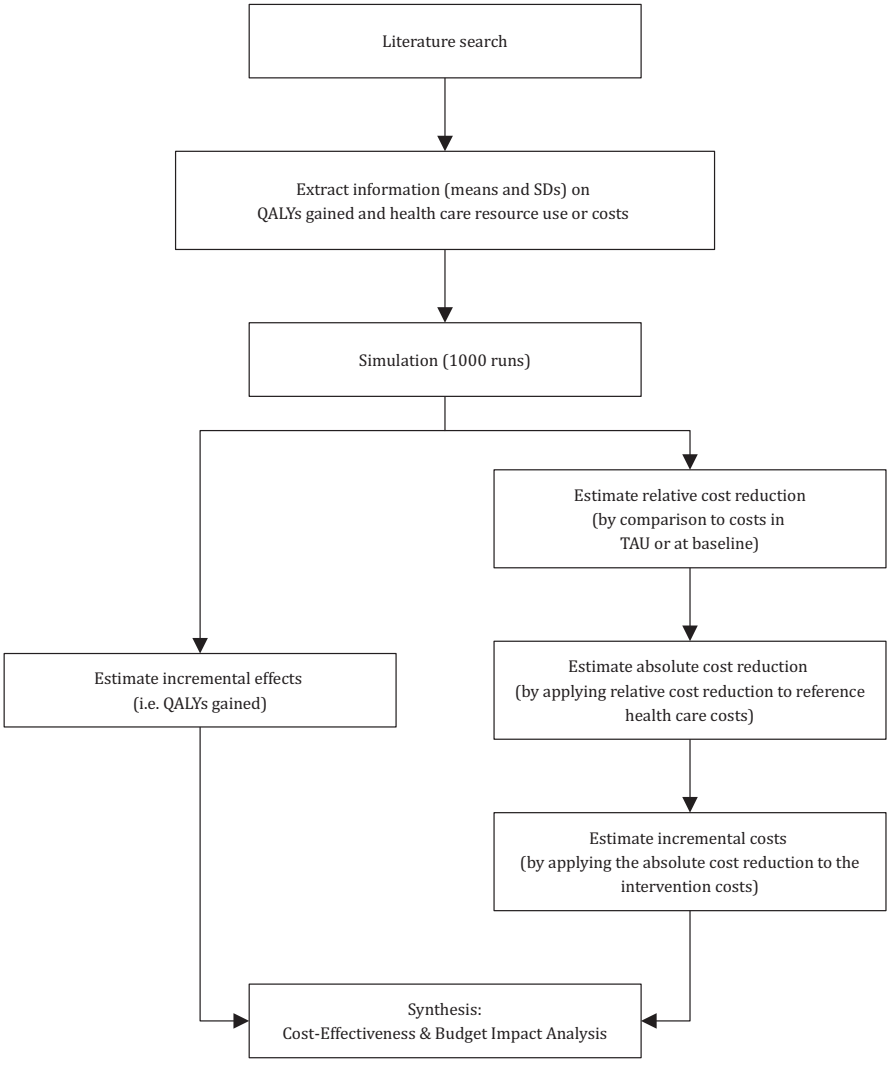
ments can be assumed to lead to improved health outcomes in addition to potential cost reductions, also their cost-effectiveness is strongly suggested.

An important question is whether this suggestion still holds when taking into account all relevant study findings that are available in the English-language literature. However, we are unaware of any method that has been described to date for synthesizing the available evidence on clinical effectiveness and health care resource use. Importantly, also studies that were not principally designed as economic evaluations provide relevant findings that should be included. These studies often investigate either clinical effectiveness or health care resource use alone. In addition, much relevant evidence has been found in non-controlled studies. The current study aims to present a method for the synthesis of all the available evidence on clinical effectiveness and health care resource use in the English literature in order to assess the cost-effectiveness and the budget impact of specialized outpatient psychotherapy for the Netherlands. The results may support policy makers in the Netherlands in deciding whether or not to allocate health care budget for the provision and upscaling of such interventions.

3.3 Methods

Through a systematic search of the English-language literature, we first identified published studies reporting on the clinical effectiveness, health care costs or resource use associated with any of the four types of specialized psychotherapies for BPD. Second, from the studies reporting on costs or resource use we extracted information regarding the relative change in health care costs following specialized psychotherapy and from the studies reporting on clinical effectiveness we extracted information on changes in health outcomes, or more specifically, the number of quality-adjusted life years (QALYs) gained. Third, the estimated relative change in health care costs was applied to the reference health care costs incurred by Dutch patients with BPD who do not receive specialized psychotherapy to calculate the reduction in health care costs, other than intervention costs, that could be expected following specialized outpatient psychotherapy. Fourth, the reduction in health care costs, other than the intervention costs, following specialized outpatient psychotherapy was compared to the costs of the intervention. Subsequently, we calculated the impact of providing an additional 1,000 patients with BPD in the Netherlands with specialized outpatient psychotherapy on the Dutch national health care budget. Fifth, a synthesis of health care costs and clinical effectiveness was performed to assess cost-effectiveness. A schematic summary of the different steps is provided as a flowchart in Figure 3.1. In an Appendix to this manuscript, we provide the mathematical details of the methodology.

Figure 3.1: Schematic summary of the methodology for this study.



3.3.1 Literature Search and Selection Criteria

Since we aimed to synthesize the available evidence on clinical effectiveness and health care resource use or costs, we scrutinized the English-language literature on any of the four types of specialized outpatient psychotherapies for relevant information. Regarding clinical effectiveness, we included studies that report changes in quality of life that were directly measured using the EuroQol-5D (EQ-5D) as well as studies that report changes in depressive symptoms that were measured using the Beck Depression Inventory (BDI). As explained in more detail in the next section, BDI scores were converted to EQ-5D scores on the basis of a mapping function. Regarding health care costs, we limited our analyses to studies on health care costs or health care resource use, as well as studies on the number of inpatient days without reporting other health care use. The last category, although providing an incomplete picture, was included, as inpatient days were considered a major cost driver in patients with BPD (Bateman & Fonagy, 2003; Wagner et al., 2014). To facilitate comparability between studies, we only included studies primarily aimed at investigating outpatient psychotherapy for BPD. Furthermore, we only included studies in which the patient sample consisted of adults. We searched for relevant literature by checking the references in a Cochrane review on psychotherapy for BPD (Stoffers et al., 2012) and a meta-analysis on DBT for BPD (Kliem et al., 2010). In addition, we performed a search in PubMed using the search terms [schema OR transference OR mentali* OR dialectical AND borderline AND therapy AND effectiveness]. A search was also performed in the NHS Economic Evaluation Database using the search term [personality disorder OR personality disorders]. The search was deliberately kept broad in order to retrieve all relevant studies. Finally, members of the Guideline Development Group for the development of the Dutch multidisciplinary guideline for borderline personality disorder could contribute studies they deemed relevant for answering the research question. The results of the literature search were reported in a flowchart (Figure 3.2), in accordance with the PRISMA statement (Moher et al., 2009).

3.3.2 Information on Health Care Costs and Quality of Life

From the included studies, the mean health care costs (i.e., excluding the costs for outpatient psychotherapy) or resource use over the course of one year of psychotherapy were compared to the mean health care costs or resource use during treatment as usual (TAU). Research designs that include TAU as a comparator condition are considered the 'gold standard' in cost-effectiveness research as it allows an assessment of whether experimental treatments have added value for current practice or not (Gold et al., 1996). Otherwise, for trials that did not include TAU as a comparator condition, we compared

the mean health care costs or resource use over the course of one year of psychotherapy to the health care costs before psychotherapy commenced (i.e., costs or resource use at baseline). We expressed changes in health care costs as percentages, to take into account the fact that pre-treatment/ baseline costs or resource use may differ between studies and countries, thereby enabling the pooling of results from studies using either costs or resource use as outcomes.

For each of the included studies that provided relevant information we also calculated the average improved health outcomes following specialized outpatient psychotherapy. To inform policy makers who decide on health care resource allocation, improved health outcomes need to be measured generically so that they can be compared across various diagnoses. Most often, quality-adjusted life years (QALYs) are used for this purpose in health economic evaluation studies. Therefore, we calculated the average number of QALYs gained of specialized outpatient psychotherapy over the course of one year (for studies in which a comparison was made with TAU, we subtracted the QALYs gained in TAU from those in the experimental condition). For studies that used the EQ-5D this could be done directly based on EQ-5D scores, for studies that used the BDI instead, the BDI scores were mapped to EQ-5D scores following the mapping function estimated by Brazier et al., 2006: EQ-5D score = 1.11 - 0.021*BDI score.

It was anticipated that only few studies report relevant information for a follow-up time beyond 12 months. Therefore, we considered a time horizon of 12 months. For the included studies that reported the relevant information only for time points earlier than or beyond 12 months, changes in clinical effectiveness and health care resource use were linearly extrapolated or interpolated, respectively.

3.3.3 Intervention Costs

Intervention costs were based on the costs of DBT, because this form of psychotherapy was investigated in most of the included studies. Moreover, we consider this a conservative approach because, in comparison to the other forms of specialized psychotherapy, DBT comprises several components in addition to the individual sessions (and therapists' consultation meetings). DBT consists of four elements: individual sessions, group sessions for skills training, telephone availability of the therapist and therapists' consultation meetings. We calculated the weighted (by sample size) average of the number of individual and group sessions attended reported in studies in which DBT was delivered in its original form according to the manual of Linehan (Linehan, 1993). Since it is the number of sessions actually attended that brings about clinical effectiveness, the calculation of the real-world intervention costs was based on this number and not on the number of sessions prescribed per protocol. We also calculated average costs for telephone availability

and consultation meetings based on studies reporting this information.

3.3.4 Cost Valuation

All costs, including the reference health care costs (i.e., costs incurred over a 12 month time period by Dutch patients with BPD who do not receive specialized psychotherapy), were indexed to 2015 euros, by using consumer price indices from the Dutch bureau of statistics. The Dutch standard cost prices (Hakkaart-van Roijen et al., 2015) were used for the valuation of health care resource use in studies that reported the use of other health care resources in addition to inpatient days, and for the calculation of the intervention costs.

3.3.5 Simulation

For each of the included studies we simulated patient-level data for both the relative cost reductions and QALYs gained. The mean and standard deviation of the health care costs or resource use over a period of 12 months of psychotherapy in each study served as inputs for the parameters of gamma distributions. From these distributions, random draws were taken in a number equal to the original study sample size. To estimate the relative cost reductions, these values were compared with the health care costs or resource use over a period of 12 months of TAU when available or pretreatment/ baseline costs otherwise. The mean and standard deviation of the number of QALYs gained over 12 months in each study served as input parameters for normal distributions. From these distributions, random draws were taken in a number equal to the original study sample size. For each simulation run, an overall weighted (by sample size) mean relative cost reduction and an overall weighted mean number of incremental QALYs were calculated. Finally, in a probabilistic sensitivity analysis the simulation was repeatedly performed for 1,000 simulation runs to estimate the confidence intervals for the mean relative cost reductions and the mean incremental QALYs.

3.3.6 Cost-effectiveness and Budget Impact

The estimated relative reduction in health care costs, apart from intervention costs, was applied to the 12 months' reference health care costs of patients with BPD in the Netherlands who do not receive specialized outpatient psychotherapy. Subsequently, by subtracting the absolute reduction in costs in the Netherlands from the additional costs of providing the interventions, the incremental costs of providing an individual Dutch BPD patient with specialized outpatient psychotherapy were calculated.

To assess the cost-effectiveness of specialized outpatient psychotherapy, the incremental costs and incremental effects (i.e., the number of QALYs gained) were synthesized. For each pair of costs and effects (CE-pair) from the 1,000 simulation runs, the costs per QALY gained were compared with the willingness-to-pay for one QALY by calculating the net monetary benefits (Hoch et al., 2002). If on average a QALY is gained for the same amount as or less than the willingness-to-pay for one QALY, then the investment can be considered cost-effective. The probability of cost-effectiveness is shown in a cost-effectiveness acceptability curve (CEAC), presenting the willingness-to-pay values on the x-axis and the probability of cost-effectiveness on the y-axis (Van Hout et al., 1994). The estimates for both the number of QALYs gained and the incremental costs resulting from each simulation run are displayed in a cost-effectiveness plane (CE-plane), yielding a cloud of 1,000 costs-effects (CE) pairs.

To assess the budget impact of scaling up the intervention to a national level, adequate information is needed on (i) the total number of patients with BPD in the Netherlands, (ii) the proportion of Dutch patients with BPD who seek help, (iii) the proportion of Dutch patients with BPD already receiving specialized outpatient psychotherapy, and (iv) the proportion of Dutch patients with BPD eligible for the intervention (e.g., see Mauskopf et al., 2007 and Sullivan et al., 2014). Unfortunately, information on these aspects is not available to our knowledge. For illustrative purposes, we calculated the budget impact of providing specialized outpatient psychotherapy to an additional 1,000 patients with BPD in the Netherlands based on the incremental costs. Although the exact number of additional patients with BPD to be treated with specialized outpatient psychotherapy after upscaling remains to be determined, we still considered it useful to perform the budget impact analysis following the abovementioned strategy and using the inputs that are available in order to demonstrate the feasibility of our approach.

3.3.7 Cost Perspective

For both the cost-effectiveness and the budget impact analysis, a health and social care system perspective was used. Hence, only health care (including social services) resource use was considered whereas other (societal) costs (e.g., productivity losses and informal care) were not. When evidence is synthesized for the specific purpose of performing a budget impact analysis, the health and social care system perspective is indeed recommended (Mauskopf et al., 2007). However, for a cost-effectiveness analysis a societal perspective is recommended. We anticipated that the availability of literature on studies reporting societal costs is limited (Wetzelaer et al., 2016). Therefore, also for the cost-effectiveness analyses, a health and social care system perspective was used. The results of our cost-effectiveness analyses are therefore limited to the health and social care system

perspective.

3.3.8 Sensitivity Analyses

In addition to the probabilistic sensitivity analysis described, we performed the following one-way deterministic sensitivity analyses to assess the influence of several uncertain components and assumptions in our calculation on the results obtained: analyses using either the lowest and highest estimate for the reference health care costs, an analysis for which only those studies that compare specialized outpatient psychotherapy to TAU were included (to address issues concerning the internal validity of uncontrolled studies), an analysis including only studies that focused on DBT (to assess the extent to which findings for DBT are comparable to the other psychotherapies), an analysis based on a more conservative estimation of QALYs gained (the number of QALYs gained in each study is reduced by 30%), and analyses that assumed either a reduction or an increase in the intervention costs by 25%.

3.4 Results

3.4.1 Literature Search

As shown in Figure 3.2, the literature search yielded 243 records in total. Based on the selection criteria, 22 studies were included. After full-text screening, 7 studies were excluded because no EQ-5D, BDI or number of inpatient days were assessed. Also, 2 studies that did use the BDI were excluded, because no BDI scores were reported.

3.4.2 Health Care Costs

The included studies that provide information on costs or resource use are listed in Table 3.1. For each study the sample size is indicated, as well as which type of psychotherapy was investigated, whether the calculation of the relative health care costs is based on a comparison with TAU or a pre-post comparison and the percentage of health care costs, excluding the intervention costs, relative to either 12 months of TAU or the 12 months preceding the psychotherapy, incurred over the course of 12 months.

Figure 3.2: Flowchart of the Literature Search.

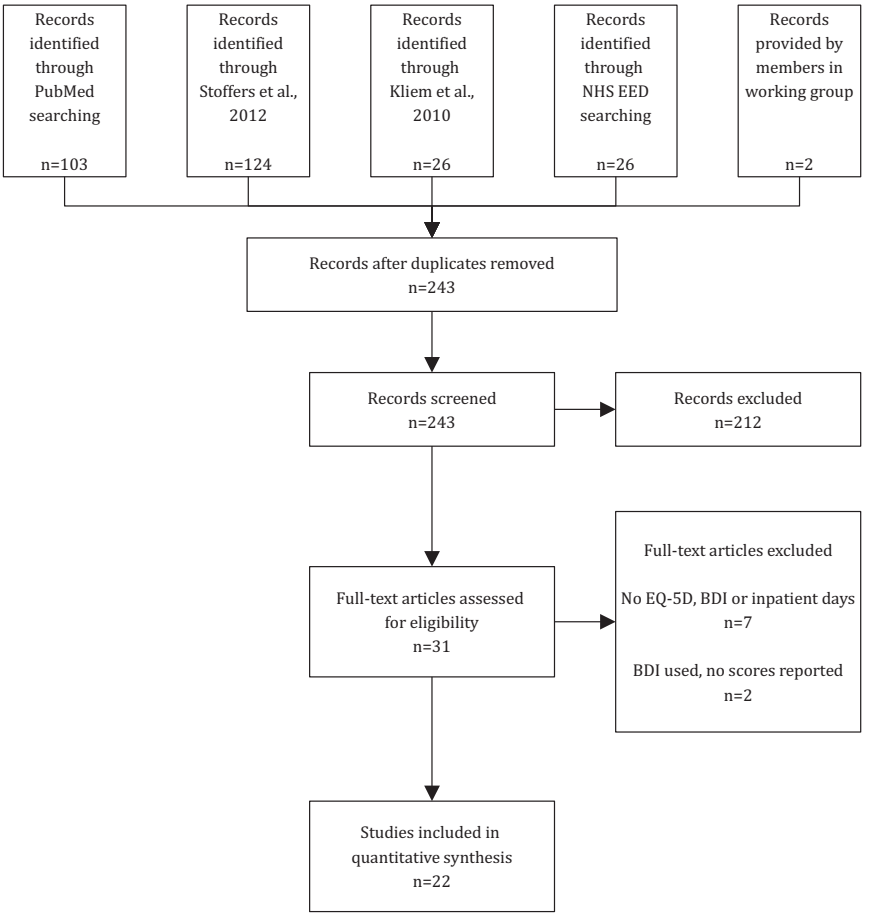


Table 3.1: Studies that Provided Information on Costs or Resource Use, Their Characteristics and the Relative Change in Health Care Costs.

Study	N	Type	Comparison	% Health care costs relative to TAU or baseline
Heard, 2000	22	DBT	TAU	22.52
Turner, 2000	12	DBT-oriented	Pre-Post	16.76
Clarkin et al., 2001	23	TFP	Pre-Post	53.68
Brassington & Krawitz, 2006	10	DBT	Pre-Post	70.59[1]
Comtois et al., 2007	23	DBT	Pre-Post	33.45
Prendergast & McCausland, 2007	11	DBT	Pre-Post	56.81[1]
van Asselt et al., 2008	44	ST	Pre-Post	37.66
van Asselt et al., 2008	42	TFP	Pre-Post	44.44
Bateman & Fonagy, 2009	71	MBT	Pre-Post	9.06
McMain et al., 2009	90	DBT	Pre-Post	28.86
Carter et al., 2010	38	DBT	TAU + waiting list	107.59[1]
Doering et al., 2010	52	TFP	TAU[2]	44.05
Pasieczny & Connor, 2011	40	DBT	TAU + waiting list	45.36[1]
Feigenbaum et al., 2012	25	DBT	TAU	80.40
Nadort, 2012	62	ST	Pre-Post	29.21[3]
Priebe et al., 2012	40	DBT	TAU	67.29
Stiglmayr et al., 2014	47	DBT	Pre-Post	4.42
Wagner et al., 2014	47	DBT	Pre-Post	26.22

Abbreviations: DBT = dialectical behaviour therapy, ST = schema therapy, TFP = transference focused psychotherapy, MBT = mentalization based treatment, TAU = treatment as usual.

Notes:

[1] These studies were conducted over a six months' time period, the relative change in health care costs was therefore multiplied by two to estimate the expected change over 12 months.

[2] In this study, TAU was given by experienced community psychotherapists.

[3] For this study only the healthcare costs incurred over a three year follow-up time period were reported, these were therefore divided by three to estimate the expected change over 12 months.

In total, the included studies represent a sample of 699 patients that provide information on health care utilisation. On average (weighted by sample size) the health care costs are reduced to 38.54% (median = 29.21%) of those in TAU or at baseline.

3.4.3 Quality of Life

The included studies that provide information on changes in quality of life (i.e., number of QALYs gained) following 12 months of specialized psychotherapy are presented in Table 3.2. Only two of the included studies made use of the EQ-5D to measure utilities, which were subsequently used to calculate QALYs. In the other studies, changes in quality of life could be estimated by mapping scores on the BDI to EQ- 5D scores.

Table 3.2: Studies that Provide Information on Changes in Quality of Life, Their Characteristics and the Mean Number of QALYs Gained.

Study	N	Type	Instrument	Mean QALYs gained (SD)
Turner, 2000	12	DBT-oriented	BDI	0.17 (0.94)
Koons et al., 2001	10	DBT	BDI	-0.04 (0.83)[1,2]
Clarkin et al., 2007	23	TFP	BDI	0.01 (0.07)[3]
Clarkin et al., 2007	23	DBT	BDI	0.01 (0.07)[3]
Prendergast & McCausland, 2007	11	DBT	BDI	0.10 (0.82)[2]
Stanley et al., 2007	20	DBT	BDI	0.15 (1.03)[4]
van Asselt et al., 2008	44	ST	EQ-5D	0.03 (0.32)
van Asselt et al., 2008	42	TFP	EQ-5D	0.09 (0.29)
Bateman & Fonagy, 2009	71	MBT	BDI	0.09 (0.91)
McMain et al., 2009	90	DBT	BDI	0.20 (0.78)
Nadort, 2012	62	ST	EQ-5D	0.13 (0.30)
Doering et al., 2010	52	TFP	BDI	0.00 (0.83)[1]
Stiglmayr et al., 2014	47	DBT	BDI	0.11 (0.87)
Gregory & Sachdeva, 2016	25	DBT	BDI	0.08 (0.84)

Continued on next page

Table 3.2 – continued from previous page

Abbreviations: DBT = dialectical behaviour therapy, ST = schema therapy, TFP = transference focused psychotherapy, MBT = mentalization based treatment, SD = standard deviation, BDI = Beck depression inventory, EQ-5D = EuroQol-5D (5 dimensions).

Notes:

[1] In these studies a comparison was made with TAU, therefore the number of QALYs gained in TAU were subtracted from those in the experimental condition.

[2] These studies were conducted over a six months' time period, the number of QALYs gained was therefore multiplied by two to estimate the incremental effects over 12 months.

[3] SDs for this study were conservatively estimated based upon the highest (excluding Doering et al., 2010 as the mean: SD ratio was very high in this study due to the mean being close to 0) mean: SD ratio found in the other studies.

[4] In this study of six months DBT BDI scores were only available for the assessment at three months, the number of QALYs gained was therefore multiplied by four to estimate the improvements in quality of life over 12 months.

In total, the included studies represent a sample of 526 patients that provide information on changes in quality of life. The weighted average number of QALYs gained is 0.08 (median = 0.09).

3.4.4 Reference Health Care Costs

The sources that were used to estimate the reference health care costs over a 12 month time period of Dutch patients with BPD who do not receive specialized psychotherapy are listed in Table 3.3. For each source the mean health care costs are listed, expressed over a time period of 12 months during which specialized outpatient psychotherapy was not systematically provided, as well as the corresponding sample sizes. These sources include studies that were identified from our literature search (van Asselt et al., 2008; Nadort, 2012) or suggested by members of the working group for the development of the Dutch multidisciplinary guideline for personality disorders (Bamelis et al., 2015; Laurenssen et al., 2016; Soeteman et al., 2008). Additionally the Vektis database was used which contains health care resource use data that is provided by all Dutch health insurers as well as other parties. The Vektis data represent the average health care costs of adult patients with personality disorders (who are diagnosed as such and whose health care costs exceed zero) in the Netherlands. Also included is one study on schema therapy for personality disorders other than BPD to partly compensate for the small number of studies that provide information on this aspect (Bamelis et al., 2015). The reported costs from this study are well in line with the other studies. The weighted (by sample size) average costs equal € 6319.

Table 3.3: Sources Used to Estimate the Average 12 Month Reference Health Care Costs of Dutch Patients with BPD.

Source	N	Mean 12 Months' reference health care costs (in 2015 euros)
van Asselt et al., 2008	86	5,913[1]
Soeteman et al., 2008	1,740	10,264[2]
Nadort, 2012	62	4,878
Vektis, 2013	65,718	6,144
Bamelis et al., 2015	320	12,082
Laurenssen et al., 2016	403	13,492

Notes:

[1] Based on a re-analysis of the data in which only health care costs were taken into account.

[2] Soeteman et al., 2008 report the increase in societal costs specific for BPD in comparison to other personality disorders (PDs). Furthermore, they report that 66.5% of societal costs are health care costs. Therefore, 66.5% of the increase in societal costs for BPD has been added to the average health care costs for all PDs to calculate the average health care costs for BPD.

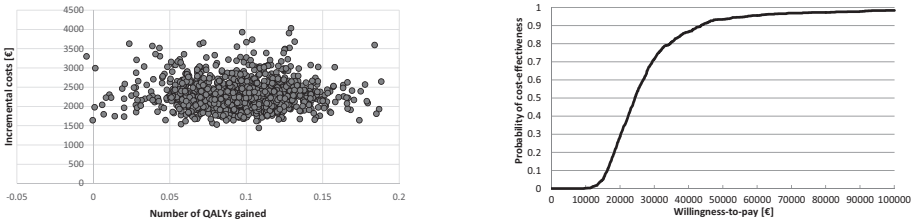
3.4.5 Intervention Costs

Two studies on DBT, delivered according to the manual of Linehan (Linehan, 1993), were used for estimating the intervention costs. In the study by McMain et al., 2009 (n=90), patients attended on average 32 individual sessions and 26 group sessions and in the study by Wagner et al., 2014 (n=47) patients attended on average 33.7 individual sessions and 16.9 group sessions. The weighted averages (by sample size) are 32.5 individual sessions and 23 group sessions. Although only little information is available on the costs for providing telephone availability of the therapist and the frequency of therapists' consultation meetings, we assumed that, similar to Wagner et al., 2014, the costs of these two components equal 26% of the costs for individual and group sessions combined. In total, intervention costs are estimated to be € 6,249.

3.4.6 Cost-effectiveness and Budget Impact

Given an average reduction in health care costs to 38.54% (95% C.I.: 28.27 - 52.88%) of the reference health care costs, when applied to the average reference health care costs of € 6,316, the expected average cost saving following specialized outpatient psychotherapy

Figure 3.3: Cost-effectiveness Plane (*left*) and Cost-effectiveness Acceptability Curve (*right*) of the Main Analysis.



Notes: The figures are based on 1,000 simulations. In the CE-plane, each dot represents a CE-pair from one simulation.

is € 3,882. Subsequently, the incremental costs for providing such treatment can be calculated by subtracting the reduction in costs from the intervention costs (€ 6,249), which gives € 2,367 (95% C.I.: € 1717 - 3272). Since the weighted average number of QALYs gained is 0.08 (95% C.I.: 0.03 - 0.16), the incremental cost-effectiveness ratio (ICER) for specialized outpatient psychotherapy is € 29,588 (95% C.I. € 13,455 - 75,940) per QALY gained. In Figure 3.3, both the CE-plane (Figure 3.3; left) and CEAC (Figure 3.3; right) of the main analysis are displayed.

At the willingness-to-pay value of € 50,000 considered acceptable for gaining one QALY by the Dutch guideline (Zwaap et al., 2015), given the burden of disease of 0.54 for BPD (Vos & Mathers, 2000), specialized outpatient psychotherapy for BPD has a 94% probability of being cost-effective. To calculate the budget impact of providing an additional 1,000 patients in the Netherlands with specialized outpatient psychotherapy, the incremental costs are multiplied by 1,000, which results in € 2,367,000 (95% C.I.: € 1,717,000 - 3,272,000).

3.4.7 Sensitivity Analyses

In Figure 3.4, the CEACs from the various sensitivity analyses are presented. In the first sensitivity analysis (Figure 3.4), instead of a weighted average of estimates for the reference health care costs of Dutch patients with BPD who do not receive specialized outpatient psychotherapy we used the lowest estimate. This decreased cost-effectiveness, so that specialized outpatient psychotherapy has a 83% probability of being cost-effective at a willingness-to-pay value of € 50,000. In a sensitivity analysis in which we used the highest estimate found in the literature for the reference health care costs (not shown in

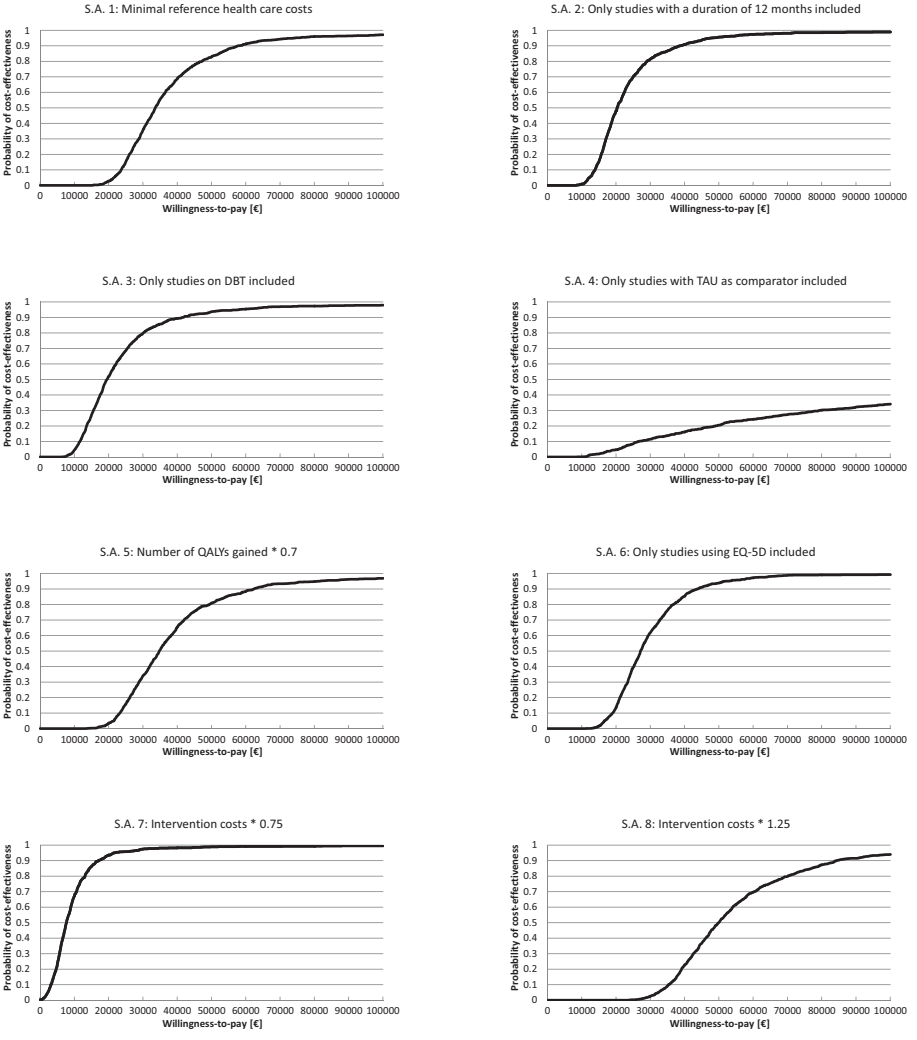
Figure 3.4), cost-effectiveness is increased to the extent that it has a 98-100% probability of cost-effectiveness over the whole range of willingness-to-pay values. In the second sensitivity analysis (Figure 3.4), in which we only included studies with a 12 month duration, there was a 96% probability of cost-effectiveness at a willingness-to-pay value of € 50,000. The third sensitivity analysis (Figure 3.4), for which only studies on DBT were included, produced results that are comparable to the main analysis, with a 94% probability of cost-effectiveness at a willingness-to-pay value of € 50,000. The fourth sensitivity analysis (Figure 3.4), in which only studies using TAU as the comparator condition were included, demonstrates that the results are sensitive to the choice regarding the inclusion of studies based on the comparator condition that was used. This analysis shows that the probability of cost-effectiveness is then decreased to 21%. When the number of QALYs gained in the included studies was conservatively assumed to be only 70% of what is originally reported, as in the fifth sensitivity analysis (Figure 3.4), cost-effectiveness is reduced. At a willingness-to-pay value of € 50,000 the probability of cost-effectiveness is 81%. In the sixth sensitivity analysis (Figure 3.4), that only included studies that used the EQ-5D, there is a 94% probability of cost-effectiveness at a willingness-to-pay value of € 50,000. When the costs of the intervention were assumed to be 25% lower than estimated in the main analysis, cost-effectiveness is increased, as shown in the seventh sensitivity analysis (Figure 3.4), with a 99% probability of cost-effectiveness at a willingness-to-pay value of € 50,000. Conversely, the eighth sensitivity analysis (Figure 3.4) shows that increasing intervention costs by 25% decreased the cost-effectiveness so that the probability is 50% at a willingness-to-pay value of € 50,000.

3.5 Discussion

3.5.1 Main Findings

By synthesizing the available evidence on the clinical effectiveness and health care resource use following 12 months of specialized outpatient psychotherapy, our results suggest that these treatments provide good value for money. We have demonstrated that, based on the available evidence, providing specialized outpatient psychotherapy to an additional 1,000 Dutch patients with BPD requires an investment of near € 2.4 million. Taking into account the improved health outcomes following these treatments, the investment needed for up-scaling can be considered cost-effective. The sensitivity analyses demonstrate that our findings are robust to most, but not all, of the alternative approaches to the methodological choices and assumptions that were made.

Figure 3.4: Cost-effectiveness Acceptability Curves of the Sensitivity Analyses.



Notes: From left to right, top to bottom: S.A. 1 = lowest estimate used for reference health care costs; S.A. 2 = only studies with a duration of 12 months were included; S.A.3 = Only studies on DBT were included; S.A. 4 = Only studies with TAU as comparator condition were included; S.A. 5 = number of QALYs gained multiplied by a half for all included studies regarding clinical effectiveness; S.A. 6 = Only studies that used the EQ-5D were included regarding clinical effectiveness; S.A. 7 = 25% reduction in intervention costs assumed; S.A. 8 = 25% increase in intervention costs assumed.

3.5.2 Strengths and Limitations

In this article, we have presented a method to assess the cost-effectiveness and budget impact of specialized outpatient psychotherapy for borderline personality disorder (BPD) in the Netherlands. The strength of this methodology is that a wealth of information has been incorporated and synthesized to inform policy making in the Netherlands. It has the advantage of incorporating various sources of empirical evidence: from both the clinical as well as the health economic research fields, from study designs that include, but are not limited to, RCTs, and from studies conducted in different countries. It further presents a way of pooling the parameter uncertainty found in the results from various studies. From the evidence available in the English-language literature, we estimated the overall average relative reduction in health care costs and applied this to the average reference health care costs of patients with BPD in the Netherlands. The incremental costs were calculated by subtracting the estimated reduction in health care costs from the intervention costs.We also estimated the incremental effects as the overall average number of QALYs gained based on the available evidence in the international literature. The incremental costs and incremental effects were synthesized to assess the cost-effectiveness and the incremental costs were further used to analyse the budget impact for the up-scaling of specialized outpatient psychotherapy for BPD in the Netherlands.

This study also has its limitations. To assess reductions in health care resource use following specialized outpatient psychotherapy we have taken into account studies that as a minimum reported on the number of inpatient days, because these are assumed to be major drivers of health care costs in patients with BPD (Bateman & Fonagy, 2003; Wagner et al., 2014). Since this type of information is often reported in articles in which the main focus is on clinical effectiveness, an advantage of our approach is that more studies could be included than if we had only included studies reporting a full health care cost analysis or health economic evaluation. However, we have also included other health care resource use where possible. It can be argued that the extent to which inpatient days are indeed major drivers of health care costs may differ between countries. Furthermore, the extent to which reductions in inpatient days are also exemplary for reductions in the use of other health care resources has not been currently investigated.

Ideally, studies on outpatient psychotherapy are designed to have an adequate follow-up time period to reveal the full impact of the intervention, including both clinical as well as health economic parameters. Unfortunately, this is not the case for many of the studies on psychotherapy for BPD. For this reason, the time horizon for the analysis performed in this article has been limited to one year. For the studies that were included that only considered a six month time period findings were linearly extrapolated to one year, thus assuming that no ceiling effects occurred within the next six months of time.

For the study by Nadort (2012), which had a duration of three years, total health care costs were divided by three, thus assuming that for this study costs increased linearly over time. A sensitivity analysis was performed to assess the influence of these assumptions, which only included studies with a duration of 12 months. The results of this analysis were comparable to the main analysis, thereby demonstrating the robustness of the latter.

Another limitation is that only two of the included studies directly measured changes in quality of life using the EQ-5D. Therefore, also studies were included that used the BDI as a measure for depressive symptoms, as BDI scores could be converted to EQ-5D scores by using a mapping function. Obviously, there is not a one-to-one correspondence between the impact of depressive symptoms and BPD symptoms on quality of life, thereby making the validity of this conversion procedure questionable. Nonetheless, owing in particular to its focus on the subjective experiences and cognitions in depression, which also play an important role in BPD (although they are qualitatively different than in Major Depressive Disorder), the BDI is considered to be sensitive in patients with BPD (Kohling et al., 2015; Silk, 2010). In a sensitivity analysis, we included only those studies that measured changes in quality of life directly using the EQ-5D. The results of this analysis are well in line with those from the main analysis. In another sensitivity analysis, we reduced the number of QALYs gained in each study by 30%. Even with this conservative approach, specialized outpatient psychotherapy still has an 81% probability of cost-effectiveness for a willingness-topay value of € 50,000, which is the threshold for disorders with a burden of disease similar to BPD in the Netherlands.

A further drawback is that not all of the included studies were RCTs with TAU as a comparator to the experimental condition. The advantage of RCTs is that they have high internal validity, so that differences between treatment conditions can be causally attributed to the actual treatments that patients received. The choice for TAU as an appropriate comparator condition is motivated by the need for establishing whether the treatment under investigation has added value for the current practice or not. To assess the extent to which the inclusion of studies that were not RCTs or did not use TAU as the comparator condition, we have performed a sensitivity analysis that only included studies that did have TAU as a comparator condition. In contrast to the main analysis, this analysis does not support the cost-effectiveness of providing a larger number of patients with BPD with specialized outpatient psychotherapy. However, a number of factors should be considered that could explain the discrepancies between this sensitivity analysis and the main analysis, in addition to the above-mentioned concerns relating to internal validity. First, there is only very little evidence available from studies comparing specialized outpatient psychotherapy for BPD with TAU: only two studies regarding the clinical effectiveness (representing only 62 participants in total) and six studies regarding health care resource use were included that used TAU as a comparator condition. Sec-

ond, both studies on the clinical effectiveness were based on the BDI (similar to most of the included studies; see above). Therefore, to gain better insight into the sensitivity of our results to the choice of included studies based on the comparator condition used, more research is needed.

In addition to gaining better insight into how specialized outpatient psychotherapy as a whole compares to TAU, the future availability of more studies that compare a specific type of specialized outpatient psychotherapy with TAU would also open new methodological avenues. For example, (Bayesian) network meta-analyses could then be used to explore how the different types of specialized outpatient psychotherapy compare to each other. At present, we considered the available evidence too scarce for the application of this method. This relates to the fact that there is currently only little evidence available for specialized outpatient psychotherapy other than DBT and that there are only a few studies that compare a specific type of specialized outpatient psychotherapy to TAU or another specific type of specialized outpatient psychotherapy.

A cautious approach was taken to deal with the uncertainty surrounding the intervention costs. In our study, the intervention costs were based on the costs of DBT, as this form of psychotherapy is the most comprehensive one (i.e., consisting of the most different components) of all four forms of psychotherapy considered and was therefore also assumed to be the most costly. We have performed sensitivity analyses in which the intervention costs were varied both by + and - 25%. The results of the first indicate that with increased intervention costs, specialized outpatient psychotherapy has a 50% probability of cost-effectiveness for a willingness-to-pay value of € 50,000. This demonstrates that our findings are sensitive to alternative assumptions involving particularly high intervention costs. Future research on comparisons between the different types of specialized outpatient psychotherapy as well as studies that focus on the added value of specific components of psychotherapy could help to identify the most cost-effective therapeutic approach and format of delivery. When the intervention costs are assumed to be lower, cost-effectiveness is enhanced compared to the main analysis. We furthermore performed a sensitivity analysis in which only studies on DBT were included, since DBT was the type of psychotherapy that was most often investigated. The results of this analysis are similar to the main analysis, thereby demonstrating that the main analysis is robust against the inclusion of other types of specialized psychotherapy in our investigation.

Ideally, our investigation should be based on a large number of studies that have simultaneously investigated both the effectiveness and costs of specialized outpatient psychotherapy. This would also provide the empirical evidence on the basis of which the correlation between incremental costs and incremental effects can be quantified. However, given the limited availability of such studies, it was decided to include studies that have investigated either effectiveness or costs. It is thus a limitation of our study that

we did not take into account the possible correlation between incremental costs and incremental effects. As a consequence, the real-world incremental costs and incremental effects would presumably be slightly different than in our simulation. Another limitation pertains to the possible bias that may be present in combining the findings of those studies that have investigated either effectiveness or costs. This is because when only the effectiveness of psychotherapy is investigated, the possibility exists that researchers have put extra effort into maximizing the effectiveness, thereby increasing intervention costs. Vice versa, when only costs are investigated, there might be an inclination towards keeping the costs to a minimum, possibly at the expense of the effectiveness of psychotherapy.

We included several sources in our investigation to determine the reference health care costs of patients with BPD in the Netherlands who do not receive specialized outpatient psychotherapy. There was substantial variability in findings regarding these reference health care costs. Therefore, in addition to the main analysis that was based on a weighted (by sample size) average, we have also performed sensitivity analyses that were based on both the lowest and highest reference health care costs. Although cost-effectiveness logically increases as these reference costs are assumed to be higher (i.e., higher reference costs leave more room for cost reduction than lower costs), even when based on the lowest costs, specialized outpatient psychotherapy for BPD can still be considered as cost-effective.

In general, our investigation is characterized by substantial uncertainty. The primary aim of the study however, is to illustrate a method that allows us to make the most of the available evidence. By making conservative choices wherever possible and by assessing the influence of several uncertain components and assumptions in various sensitivity analyses, our findings are useful for two reasons. First, they will help to determine how, based on the evidence that is currently available, specialized outpatient psychotherapy will have an impact on the Dutch health care budget and to what extent investments in these treatments can be considered cost-effective. Second, they provide important directions for future research. For example, when the results of a sensitivity analysis deviate substantially from the main analysis for reasons that remain largely unknown (e.g., as is the case for the sensitivity analysis that only included studies using TAU as a comparator), studies could be designed to specifically address this.

An issue that is pivotal to any synthesis of international evidence for producing country-specific estimates is related to the transferability of research findings. Indeed, a limitation of this study pertains to the fact that it is unknown to what extent the findings from one country are transferable to another. As mentioned before, we assumed inpatient days to be the major drivers for health care costs in BPD, but the extent to which this holds true may differ between countries. In a wider sense, this applies to country-specific estimates of any particular kind of health care resource use. One step towards

a resolution of this transferability issue would be to try and map out the variations between different countries by performing future studies in a multinational context. Once the variations are estimated, adaptations of research findings may possibly be devised to enhance their transferability. Although in this study we estimated the cost-effectiveness and budget impact of specialized outpatient psychotherapy specifically for the Netherlands, our methodology could, quite straightforwardly, be applied to other countries as well. To make a case for other countries, both for the reference health care costs as well as for the intervention costs, country-specific data are needed.

One final aspect of uncertainty in the context of our investigation that should be mentioned is the extent to which our findings apply to longer time periods than the one year that is currently considered. The improvements in health outcomes following specialized outpatient psychotherapy may persist over time, during which no further treatment is needed. In that case, since the investment has already been made, then also the reductions in health care costs will continue to accrue over time. The cost-effectiveness of specialized outpatient psychotherapy will then be enhanced for as long as there is no additional need for receiving further treatment. However, to demonstrate such effects over longer time periods in a way similar to our current approach, more empirical evidence is needed from studies that employ a longer time horizon for follow-up. It would be particularly interesting to validate the findings from our current investigation with those from a large-sample RCT performed in the Netherlands, preferably with a long time period for follow-up, and preferably with TAU as a comparator condition.

Despite the limitations described above, we have demonstrated an approach that allows the synthesis of all relevant study results that are available, be they from clinical studies or health economic investigations, and from RCTs as well as non-controlled trials, in a way that respects the uncertainty that surrounds those findings and will help to shed light not only on the cost-effectiveness and the budget impact per se, but also on which remaining questions could be addressed in future studies. The results from our investigation may support policy decisions about the allocation of health care budget for the provision of specialized outpatient psychotherapy for borderline personality disorder patients in the Netherlands.

3.6 Acknowledgements

Pim Wetzelaer is supported by a grant from the Netherlands Organization for Health Research and Development (ZonMW; 80-82310-97-12142), which was obtained by Arnoud Arntz. We would like to thank the Guideline Development Group involved in the development of the Dutch multidisciplinary guideline for borderline personality disorder

for its valuable input, Jan Brassington for providing additional information and Annelies Laurenssen for making available important data.

3.7 Appendix

3.7.1 Incremental Effects

For each study 1, 2, ..., i, we perform j number (corresponding to the study sample size n_i) of random draws $\varepsilon_{i,1}$, $\varepsilon_{i,2}$, ..., $\varepsilon_{i,j}$ from the Normal distribution $N(\mu_i, \sigma_i^2)$, where μ_i and σ_i^2 are the mean and variance, respectively, of the number of QALYs gained during 12 months of specialized outpatient psychotherapy in study i. Subsequently, we calculate an overall, weighted by sample size, mean number of QALYs gained ΔE, following:

$$\Delta E = \frac{\sum_{i=1}^{i} \overline{e}_i * n_i}{\sum_{i=1}^{i} n_i}. \tag{3.1}$$

This procedure is replicated in 1,000 simulation runs to provide the range of values for the incremental effects E from which we calculate the mean and 95%-confidence intervals and which are synthesized with the incremental costs.

3.7.2 Incremental Costs

For each study 1, 2, ..., i, we perform j number (corresponding to the study sample size n_i) of random draws $c_{i,1}$, $c_{i,2}$, ..., $c_{i,j}$ from the Gamma distribution $\Gamma(\alpha_i, \beta_i)$, where α_i and β_i are the shape and rate parameters, respectively, for the distribution of the health care costs incurred during 12 months in study i. The values for these parameters can be calculated from the mean μ_i and standard deviation σ_i, following $\alpha_i = (\mu_i/\sigma_i)^2$ and $\beta_i = \sigma_i^2/\mu_i$. Subsequently, we calculate the overall, weighted by sample size, mean incremental costs ΔC, following:

$$\Delta C = C_I - \left(\frac{\sum_{i=1}^{i} \sum_{j=1}^{i} \frac{c_{i,j}}{C_{C,i}}}{\sum_{i=1}^{i} n_i} \right) * C_R, \tag{3.2}$$

where $C_{C,i}$ are the health care costs in TAU or at baseline used for comparison for study i, C_R are the reference health care costs, and C_I are the intervention costs. Also this procedure is replicated in the 1,000 simulation runs to provide the range of values for the incremental costs from which we calculate the mean and 95%-confidence intervals and which are synthesized with the incremental effects.

3.7.3 Synthesis

The incremental effects and incremental costs are synthesized by calculating the incremental net monetary benefits (INMBs), following $INMB = \lambda * \Delta E - \Delta C$, where λ is the willingness-to-pay value (Hoch et al., 2002). In each of a 1,000 simulation runs, one pair of incremental effects and incremental costs are synthesized in this way and by tracking the relative number of simulation runs in which $INMB > 0$ the probability of cost-effectiveness is estimated for the range of different willingness-to-pay values λ.

3.8 References

Arntz, A., & Van Genderen, H. (2011). *Schema therapy for borderline personality disorder*. John Wiley & Sons.

Bamelis, L. L., Arntz, A., Wetzelaer, P., Verdoorn, R., & Evers, S. M. (2015). Economic evaluation of schema therapy and clarification-oriented psychotherapy for personality disorders: a multicenter, randomized controlled trial. *J Clin Psychiatry*, *76*(11), e1432-40.

Bateman, A., & Fonagy, P. (2003). Health service utilization costs for borderline personality disorder patients treated with psychoanalytically oriented partial hospitalization versus general psychiatric care. *Am J Psychiatry*, *160*(1), 169-71.

Bateman, A., & Fonagy, P. (2004). Mentalization-based treatment of BPD. *Journal of personality disorders*, *18*(1), 36.

Bateman, A., & Fonagy, P. (2009). Randomized controlled trial of outpatient mentalization-based treatment versus structured clinical management for borderline personality disorder. *American Journal of Psychiatry*.

Bender, D. S., Dolan, R. T., Skodol, A. E., Sanislow, C. A., Dyck, I. R., McGlashan, T. H., … Gunderson, J. G. (2001). Treatment utilization by patients with personality disorders. *American Journal of psychiatry*, *158*(2), 295-302.

Brassington, J., & Krawitz, R. (2006). Australasian dialectical behaviour therapy pilot outcome study: effectiveness, utility and feasibility. *Australasian Psychiatry*, *14*(3), 313-318.

Brazier, J., Tumur, I., Holmes, M., Ferriter, M., Parry, G., Dent-Brown, K., & Paisley, S. (2006). Psychological therapies including dialectical behaviour therapy for borderline personality disorder: a systematic review and preliminary economic evaluation. *Health Technol Assess*, *10*(35), iii, ix-xii, 1-117.

Carter, G. L., Willcox, C. H., Lewin, T. J., Conrad, A. M., & Bendit, N. (2010). Hunter DBT project: randomized controlled trial of dialectical behaviour therapy in women with borderline personality disorder. *Australian and New Zealand Journal of Psychiatry*, *44*(2), 162-173.

Clarkin, J. F., Foelsch, P. A., & Levy, K. N. (2001). The developemnt of psychodynamic treatment for patients with borderline personality disorder: A preliminary study of behavioral change. *Journal of Personality Disorders*, *15*(6), 487.

Clarkin, J. F., & Kernberg, O. F. (2015). *Transference-focused psychotherapy for borderline personality disorder: A clinical guide*. American Psychiatric Pub.

Clarkin, J. F., Levy, K. N., Lenzenweger, M. F., & Kernberg, O. F. (2007). Evaluating three treatments for borderline personality disorder: a multiwave study. *American Journal of Psychiatry*.

Comtois, K. A., Elwood, L., Holdcraft, L. C., Smith, W. R., & Simpson, T. L. (2007). Effectiveness of dialectical behavior therapy in a community mental health center. *Cognitive and Behavioral Practice*, *14*(4), 406-414.

Doering, S., Hörz, S., Rentrop, M., Fischer-Kern, M., Schuster, P., Benecke, C., … Buchheim, P. (2010). Transference-focused psychotherapy v. treatment by community psychotherapists for borderline personality disorder: randomised controlled trial. *The British Journal of Psychiatry*, *196*(5), 389-395.

Feigenbaum, J. D., Fonagy, P., Pilling, S., Jones, A., Wildgoose, A., & Bebbington, P. E. (2012). A real-world study of the effectiveness of DBT in the UK National Health Service. *British Journal of Clinical Psychology*, *51*(2), 121-141.

Gold, M. R., Siegel, J. E., Russell, L. B., & Weinstein, M. C. (1996). *Cost-effectiveness in health and medicine*. New York: Oxford University Press.

Gregory, R. J., & Sachdeva, S. (2016). Naturalistic outcomes of evidence-based therapies for borderline personality disorder at a medical university clinic. *American Journal of Psychotherapy*, *70*(2), 167-184.

Hakkaart-van Roijen, L., Van der Linden, N., Bouwmans, C., Kanters, T., & Tan, S. (2015). *Kostenhandleiding. methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg.* Zorginstituut Nederland.

Heard, H. L. (2000). Cost-effectiveness of dialectical behavior therapy in the treatment of borderline personality disorder. *Dissertation of the University of Washington*.

Hoch, J. S., Briggs, A. H., & Willan, A. R. (2002). Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Econ*, *11*(5), 415-30.

Kliem, S., Kröger, C., & Kosfelder, J. (2010). Dialectical behavior therapy for borderline personality disorder: a meta-analysis using mixed-effects modeling. *Journal of consulting and clinical psychology*, *78*(6), 936.

Kohling, J., Ehrenthal, J. C., Levy, K. N., Schauenburg, H., & Dinger, U. (2015). Quality and severity of depression in borderline personality disorder: A systematic review and meta-analysis. *Clin Psychol Rev*, *37*, 13-25.

Koons, C. R., Robins, C. J., Tweed, J. L., Lynch, T. R., Gonzalez, A. M., Morse, J. Q., … Bastian, L. A. (2001). Efficacy of dialectical behavior therapy in women veterans with borderline personality disorder. *Behavior therapy*, *32*(2), 371-390.

Laurenssen, E. M., Eeren, H. V., Kikkert, M. J., Peen, J., Westra, D., Dekker, J. J., & Busschbach, J. J. (2016). The burden of disease in patients eligible for mentalization-based treatment (MBT): quality of life and costs. *Health Qual Life Outcomes*, *14*(1), 145.

Leichsenring, F., Leibing, E., Kruse, J., New, A. S., & Leweke, F. (2011). Borderline personality disorder. *The Lancet*, *377*(9759), 74-84.

Linehan, M. (1993). *Cognitive-behavioral treatment of borderline personality disorder*. Guilford Press.

Mauskopf, J. A., Sullivan, S. D., Annemans, L., Caro, J., Mullins, C. D., Nuijten, M., … Trueman, P. (2007). Principles of good practice for budget impact analysis: report of the ISPOR task force on good research practices–budget impact analysis. *Value Health*, *10*(5), 336-47.

McMain, S. F., Links, P. S., Gnam, W. H., Guimond, T., Cardish, R. J., Korman, L., & Streiner, D. L. (2009). A randomized trial of dialectical behavior therapy versus general psychiatric management for borderline personality disorder. *American Journal of Psychiatry*.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*, *339*, b2535.

Nadort, M. (2012). Wild at heart and weird on top. the implementation of outpatient schema therapy for borderline personality disorder in regular mental health care. *Dissertation of the VU University Amsterdam*.

Pasieczny, N., & Connor, J. (2011). The effectiveness of dialectical behaviour therapy in routine public mental health settings: An Australian controlled trial. *Behav Res Ther*, *49*(1), 4-10.

Prendergast, N., & McCausland, J. (2007). Dialectic behaviour therapy: a 12-month collaborative program in a local community setting. *Behaviour Change*, *24*(01), 25-35.

Priebe, S., Bhatti, N., Barnicot, K., Bremner, S., Gaglia, A., Katsakou, C., … Zinkler, M. (2012). Effectiveness and cost-effectiveness of dialectical behaviour therapy for self-harming patients with personality disorder: a pragmatic randomised controlled trial. *Psychother Psychosom*, *81*(6), 356-65.

Silk, K. R. (2010). The quality of depression in borderline personality disorder and the diagnostic process. *J Pers Disord*, *24*(1), 25-37.

Soeteman, D. I., Hakkaart-van Roijen, L., Verheul, R., & van Busschbach, J. J. (2008). The economic burden of personality disorders in mental health care. *Journal of Clinical Psychiatry*, *69*(2), 259-265.

Stanley, B., Brodsky, B., Nelson, J. D., & Dulit, R. (2007). Brief dialectical behavior therapy (DBT-B) for suicidal behavior and non-suicidal self injury. *Archives of Suicide Research*, *11*(4), 337-341.

Stiglmayr, C., Stecher-Mohr, J., Wagner, T., Meißner, J., Spretz, D., Steffens, C., … Schulze, J. (2014). Effectiveness of dialectic behavioral therapy in routine outpatient care: the berlin borderline study. *Borderline personality disorder and emotion dysregulation*, *1*(1), 1.

Stoffers, J. M., Völlm, B. A., Rücker, G., Timmer, A., Huband, N., & Lieb, K. (2012). Psychological therapies for people with borderline personality disorder. *The Cochrane Library*.

Sullivan, S. D., Mauskopf, J. A., Augustovski, F., Jaime Caro, J., Lee, K. M., Minchin, M., … Shau, W. Y. (2014). Budget impact analysis-principles of good practice: report of the ISPOR 2012 Budget Impact Analysis Good Practice II Task Force. *Value Health*, *17*(1), 5-14.

Turner, R. M. (2000). Naturalistic evaluation of dialectical behavior therapy-oriented treatment for borderline personality disorder. *Cognitive and Behavioral Practice*, *7*(4), 413-419.

van Asselt, A. D., Dirksen, C. D., Arntz, A., Giesen-Bloo, J. H., van Dyck, R., Spinhoven, P., … Severens, J. L. (2008). Out-patient psychotherapy for borderline personality disorder: cost-effectiveness of schema-focused therapy v. transference-focused psychotherapy. *Br J Psychiatry*, *192*(6), 450-7.

Van Hout, B. A., Al, M. J., Gordon, G. S., & Rutten, F. F. (1994). Costs, effects and c/e-ratios alongside a clinical trial. *Health economics*, *3*(5), 309-319.

Vos, T., & Mathers, C. D. (2000). The burden of mental disorders: a comparison of methods between the Australian burden of disease studies and the Global Burden of Disease study. *Bulletin of the World Health Organization*, *78*(4), 427-438.

Wagner, T., Fydrich, T., Stiglmayr, C., Marschall, P., Salize, H.-J., Renneberg, B., … Roepke, S. (2014). Societal cost-of-illness in patients with borderline personality disorder one year before, during and after dialectical behavior therapy in routine outpatient care. *Behaviour research and therapy*, *61*, 12-22.

Wetzelaer, P., Lokkerbol, J., Arntz, A., van Asselt, A., & Evers, S. (2016). [Cost-effectiveness of psychotherapy for personality disorders; a systematic review on economic evaluation studies]. *Tijdschr Psychiatr*, *58*(10), 717-727.

Widiger, T. A. (2012). *The oxford handbook of personality disorders*. Oxford University Press.

Zanarini, M. C. (2009). Psychotherapy of borderline personality disorder. *Acta Psychiatrica Scandinavica*, *120*(5), 373-377.

Zwaap, J., Knies, S., Van der Meijden, C., Staal, P., & Van der Heiden, L. (2015). *Kosteneffectiviteit in de praktijk*. Zorginstituut Nederland.

Chapter 4

Design of an international multicentre RCT on group schema therapy for borderline personality disorder

Pim Wetzelaer[1], Joan Farrell[2,3], Silvia MAA Evers[4,5], Gitta A Jacob[6], Christopher W Lee[7], Odette Brand[8], Gerard van Breukelen[9,10], Eva Fassbinder[11], Heather Fretwell[12,13], R Patrick Harper[14], Anna Lavender[15], George Lockwood[16], Ioannis A Malogiannis[17,18], Ulrich Schweiger[19], Helen Startup[15], Teresa Stevenson[20], Gerhard Zarbock[21] and Arnoud Arntz[1,22]

[1]Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands
[2]Department of Psychology, Indiana University-Purdue University Indianapolis, Indianapolis, USA
[3]Center for Borderline Personality Disorder Treatment & Research, Indianapolis, USA
[4]Department of Health Services Research, CAPHRI School of Public Health and Primary Care, Faculty of Health Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands
[5]Centre of Economic Evaluation, Trimbos Institute (Netherlands Institute of Mental Health and Addiction), Utrecht, The Netherlands
[6]Department of Clinical Psychology and Psychotherapy, Institute for Psychology, University of Freiburg, Freiburg, Germany
[7]Department of Psychology and Exercise Science, Murdoch University, Murdoch, Australia
[8]De Viersprong, The Netherlands Institute for Personality Disorders, Halsteren, The Netherlands
[9] Department of Methodology and Statistics, Faculty of Health Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands
[10]Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands.
[11]Department of Psychiatry and Psychotherapy, University of LÃijbeck, LÃijbeck, Germany.
[12]Midtown Mental Health/ Eskenazi Health, Indianapolis, USA
[13]Department of Psychiatry, Indiana University School of Medicine, Indianapolis, USA
[14]Bradford District Care Trust, Bradford, UK
[15]South London and Maudsley NHS Foundation Trust, London, UK
[16]Schema Therapy Institute Midwest, Kalamazoo, USA.
[17] 1st Department of Psychiatry, Eginition Hospital, Medical School, Athens University, Athens, Greece
[18]Greek Society of Schema Therapy, Athens, Greece
[19]Klinik für Psychiatrie und Psychotherapie, University of Lübeck, Lübeck, Germany
[20]Peel and Rockingham Kwinana Mental Health Service, Rockingham, Australia
[21]IVAH GmbH (Institute for Training in CBT), Hamburg, Germany
[22] Department of Clinical Psychology, University of Amsterdam, Amsterdam, The Netherlands

Reference: Wetzelaer, P., Farrell, J., Evers, S.M.A.A., Jacob, G.A., Lee, C., Brand, O., van Breukelen, G., Fassbinder, E., Fretwell, H., Harper, P., Lavender, A., Lockwood, G., Malogiannis, I., Schweiger, U., Startup, H., Stevenson, T., Zarbock, G., Arntz, A. (2014). Design of an international multicentre RCT on group schema therapy for borderline personality disorder. *BMC Psychiatry*, 14, 319-334.

4.1 Abstract

BACKGROUND Borderline personality disorder (BPD) is a severe and highly prevalent mental disorder. Schema therapy (ST) has been found effective in the treatment of BPD and is commonly delivered through an individual format. A group format (group schema therapy, GST) has also been developed. GST has been found to speed up and amplify the treatment effects found for individual ST. Delivery in a group format may lead to improved cost-effectiveness. An important question is how GST compares to treatment as usual (TAU) and what format for delivery of schema therapy (format A; intensive group therapy only, or format B; a combination of group and individual therapy) produces the best outcomes.

METHODS/DESIGN An international, multicentre randomized controlled trial (RCT) will be conducted with a minimum of fourteen participating centres. Each centre will recruit multiple cohorts of at least sixteen patients. GST formats as well as the orders in which they are delivered to successive cohorts will be balanced. Within countries that contribute an uneven number of sites, the orders of GST formats will be balanced within a difference of one. The RCT is designed to include a minimum of 448 patients with BPD. The primary clinical outcome measure will be BPD severity. Secondary clinical outcome measures will include measures of BPD and general psychiatric symptoms, schemas and schema modes, social functioning and quality of life. Furthermore, an economic evaluation that consists of cost-effectiveness and cost-utility analyses will be performed using a societal perspective. Lastly, additional investigations will be carried out that include an assessment of the integrity of GST, a qualitative study on patients' and therapists' experiences with GST, and studies on variables that might influence the effectiveness of GST.

DISCUSSION This trial will compare GST to TAU for patients with BPD as well as two different formats for the delivery of GST. By combining an evaluation of clinical effectiveness, an economic evaluation and additional investigations, it will contribute to an evidence-based understanding of which treatment should be offered to patients with BPD from clinical, economic, and stakeholders' perspectives.

TRIAL REGISTRATION Netherlands Trial Register NTR2392. Registered 25 June 2010.

KEYWORDS Borderline personality disorder, Group schema therapy, RCT, Economic evaluation, Cost-effectiveness

4.2 Background

Borderline personality disorder (BPD) is a common mental disorder characterised by enduring and pervasive patterns of instability in interpersonal relationships, identity, impulsivity, and affect (Leichsenring et al., 2011). The prevalence of BPD in the general population, as revealed by recent community surveys that use DSM-IIIR or DSM-IV criteria, is estimated at 0.5 to 2.7% (median = 0.7%) (Samuels, 2011). BPD prevents patients from developing their full potential and leading a fulfilling life. Consequently, many patients do not finish their education, or complete it at a suboptimal level. Similarly, many have a job beneath their capacity, or they have no job at all. BPD patients often engage in problematic relationships, self-injury, suicide attempts, and substance abuse. Furthermore, 8-10% of BPD patients end their lives prematurely due to suicide ("American Psychiatric Association: Practice guideline for the treatment of patients with borderline personality disorder", 2001).

BPD severely impairs quality of life across mental, social and physical dimensions (IsHak et al., 2013). In a Swedish study that compared quality of life between women with BPD and a normal population, it was found that women with BPD were significantly impaired in all domains, including physical, emotional, cognitive and sexual functioning (Perseius et al., 2006). Relationships with their family and partner were also found to be impaired (Perseius et al., 2006). Two Dutch studies have shown that the burden of disease for both adolescent and adult patients with various personality disorders, including BPD, is severe and their quality of life is markedly impaired (Feenstra et al., 2012; Soeteman, Verheul, & van Busschbach, 2008).

In addition to the devastating effects of BPD on the functioning of patients, it imposes a large burden on their families, friends, and society as a whole. Families and friends may face the challenging task of providing informal care (Bauer et al., 2012; Dunne & Rogers, 2013), whereas society bears the costs of a more intensive use of health services (van Asselt et al., 2007; Bender et al., 2001; Coid et al., 2009; Feenstra et al., 2012; Soeteman, Hakkaart-van Roijen, et al., 2008), productivity losses (van Asselt et al., 2007; Soeteman, Hakkaart-van Roijen, et al., 2008), and other intersectoral costs (Drost et al., 2013). In clinical settings, BPD patients are regarded as notoriously difficult to treat, leading many therapists to refrain from treating them. In the absence of robust evidence for the effectiveness of any specific medication for BPD (Feurino & Silk, 2011), psychotherapy is, currently, the most promising strategy for its treatment.

Schema therapy (ST) is delivered as an outpatient treatment with the intention of bringing about full recovery. It has proven more clinically effective than transference-focused psychotherapy (TFP) in a randomized controlled trial (RCT) comparing both treatments head-to-head (Giesen-Bloo et al., 2006). Results from the same RCT also show

that it has a high probability of being more cost-effective than TFP (van Asselt et al., 2008). It was found that ST could be transported out of the clinical trial to use in the regular Dutch healthcare setting with no loss of clinical effectiveness (Nadort et al., 2009). In these studies, all treatments led to an improvement in quality of life above and beyond recovery from symptoms.

ST can also be delivered in a group format, thus enabling a more efficient use of resources. In addition, initial findings have indicated that this format can increase the effectiveness of ST (Farrell et al., 2009). Psychotherapy groups can provide a family-like environment to patients, giving them a sense of belonging and facilitating the secure attachments needed for limited reparenting (a defining element of ST that refers to the therapist trying, within the bounds of a professional relationship, to meet a patient's unfulfilled core emotional needs). Furthermore, patients can accept the responses of peers as more 'genuine' than those of the therapist, whose responses are, at least initially, often viewed as less 'real' and more professional. An RCT on group schema therapy (GST) has demonstrated its effectiveness for the treatment of BPD (Farrell et al., 2009). Treatment time in this study was relatively short (eight months in contrast to up to three years for studies on individual ST). This suggests that GST leads to faster recovery than individual ST. However, this seminal investigation still leaves important questions unanswered. First, since this RCT was performed by the developers of GST it is unknown how effective GST is when delivered by other therapists in other centres. Second, in this RCT GST was an addition to treatment as usual (TAU) for patients who were already receiving TAU beforehand. GST has thus not been tested as an integral and stand-alone treatment. Third, this RCT was not accompanied by an economic evaluation, and hence, does not give insight into the cost-effectiveness of GST. Fourth, in this RCT TAU was very ineffective. Because TAU has improved in recent years, due to the dissemination of evidence-based treatments and recent insights from studies, GST needs to be compared to up-to-date TAU.

A second study on GST for BPD was a Dutch pilot study in which two cohorts of BPD patients (total number of 18 patients) were treated with the combination of group and individual ST (Dickhaut & Arntz, 2014). This study showed large effect sizes on a broad range of outcomes including improvements in BPD symptomatology, general psychopathological symptoms and quality of life. However, this study was uncontrolled and also did not assess cost-effectiveness (Dickhaut & Arntz, 2014). In sum, findings on the clinical effectiveness of GST from previous studies are promising yet leave open important questions that need to be answered before the further dissemination of GST is supported.

To answer these questions an international, multicentre RCT on GST for BPD will be performed. This article provides a description of the study design. The main study

objective is to compare the clinical effectiveness and cost-effectiveness of GST and TAU. The RCT involves two formats of GST, one that consists almost exclusively of GST and one that combines GST with individual ST. Both formats have a two year duration. Group and individual schema therapy are, to a large extent, considered complementary and mutually supportive. Individual sessions may have an advantage over group sessions in that the therapist is in a better position to motivate a patient for treatment, to offer extended trauma processing, and to offer a more in depth attachment. In contrast, the group sessions may provide important connection experiences that deal with fundamental issues of BPD. For example, a stronger sense of connection provided by the group may do more to counter abandonment fears and sharing common experiences among peers might add to a decrease in a patient's sense of isolation and/or defectiveness. On the one hand, combining group and individual schema therapy may offer potentially synergistic effects (Dickhaut & Arntz, 2014). On the other hand, the availability of individual sessions might lead to some patients avoiding full participation in the group, thus reducing its potential curative power. Hence, no specific hypothesis about the superiority of either format has been formulated. To evaluate the relative contribution of the proposed formats to outcome, a secondary objective is to compare the two formats of GST. This will help to establish the optimal format for delivery of GST to patients with BPD.

Furthermore, a series of additional substudies will be performed. These consist of an assessment of the integrity of GST, a qualitative investigation into the experiences of patients and therapists with GST and an investigation of variables that might influence the change process of GST and thereby affect outcomes such as dropout rates and patient improvement. Qualitative data will be collected from patient and therapist interviews and/or focus groups regarding their experiences of specific aspects of GST. This will provide information on which aspects of the GST protocol are most beneficial and any aspects that may be less helpful or problematic. This will not only help in identifying how the different components of ST can affect outcome, but also in deciding which format is preferred. Based on this information, GST can then be further tailored to the needs of the primary stakeholders before its further dissemination.

4.3 Methods/Design

In this RCT an evaluation of clinical effectiveness, a full economic evaluation and a series of additional investigations will be performed. Primarily, GST (format A or B) will be compared to TAU and, secondarily, GST formats A and B will be compared.

4.4 Design

A multicentre RCT will be conducted with participating centres (at the time of this writing) in the following countries: six centres in the Netherlands, three in Germany, one in Australia, two in the UK, one in the USA and one in Greece. Some centres that initially agreed to participate had to withdraw because of budget cuts resulting from economic difficulties that made participation impossible. One Dutch centre withdrew due to recruitment rates being too slow and was replaced by another centre. In contrast, two Dutch centres' expeditious recruitment rates permitted the inclusion of a third cohort. This can compensate for additional centres that agreed to participate but may still withdraw, or for centres that fail to recruit the planned minimum of 32 patients per centre. Patient flow, screening, randomization and assessments are displayed schematically in Figure 4.1. The research protocol has been approved by the Medical Ethics Committee of Maastricht University for the Dutch sites; by the Murdoch University Human Research Ethics Committee for the Australian sites; by the Ethics Committee of the Albert-Ludwigs-University Freiburg, the Ethics Committee of the University of Lübeck and the Ethics Committee of the Psychotherapist Association Hamburg for the German sites; by the Ethics Committee of the Eginition Hospital, Medical School, University of Athens for the Greek site; and by the National Research Ethics Service Committee London - Camberwell St Giles for the British sites. Ethical review is in process in the USA.

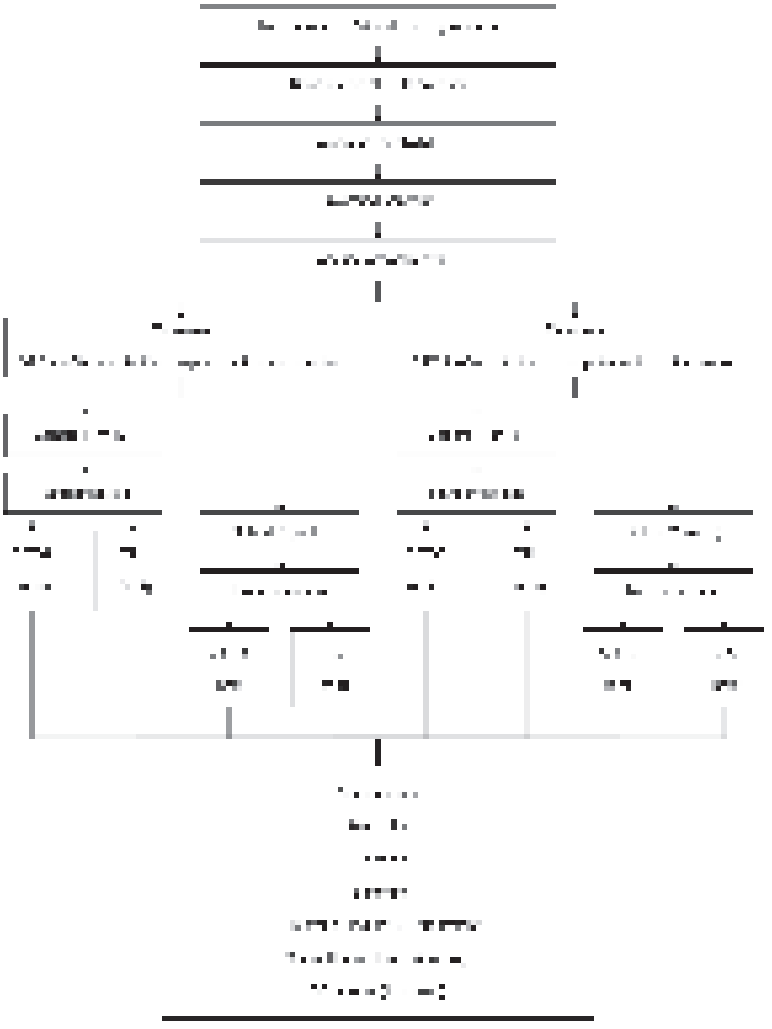
4.4.1 Recruitment

Patients will be recruited in the participating centres. They will be invited to participate in the screening process when diagnosed with BPD or when this is suspected. Both patients who are already receiving treatment for BPD as well as new referrals can be included as participants. After reading and hearing information on the RCT and signing an informed consent, patients will be assessed for in- and exclusion criteria.

4.4.2 Patients

Patients are eligible when they (1) are between 18 and 65 years of age, (2) have a primary diagnosis of BPD, (3) have a BPD severity score of above 20 on the Borderline Personality Disorder Severity Index, version IV (Arntz et al., 2003), (4) are willing to participate in the study and (5) are able to participate in (group) treatment and research for 2 years. Patients will be excluded if they have a lifetime psychotic disorder (except for a brief psychotic disorder as described in the Diagnostic and Statistical Manual of mental disorders,

Figure 4.1: Flow chart of the study design.



Notes: Patients with BPD are recruited at 14 participating centres and screened for eligibility. After informed consent is signed, baseline assessments are performed. Subsequently, patients are randomized in blocks of two per centre to either GST or TAU. Half of the centres offer GST-A to the first cohort of patients and the other half offers GST-B to the first cohort of patients. In two Dutch sites, a third cohort is recruited which is randomly assigned to either format for GST so that the total number of cohorts receiving both formats is balanced. Assessments are performed approximately every six months for the first two years, after which GST treatment ends. Costs are also assessed at approximately 30 months. Follow-up assessments take place 36 months after randomization.

version IV (DSM-IV) BPD criterion 9); an IQ below 80; if they are unable to read, speak or write the language used at the centre; if they have Attention Deficit Hyperactivity Disorder (ADHD) (if successfully treated, ADHD is not an exclusion criterion), bipolar disorder type 1, dissociative identity disorder, full or subthreshold narcissistic or antisocial personality disorder (PD), substance dependence needing clinical detoxification, a serious and/or unstable medical illness or if they have received schema therapy of more than three months duration in the last three years. Well-trained clinicians will diagnose patients at baseline using Structured Clinical Interviews for DSM-IV axis I and II disorders (SCID I and II). Screening for ADHD will be performed using a six item version of the World Health Organization (WHO) ADHD screener (Kessler et al., 2005). If this screener indicates the presence of ADHD, then the patient will be further assessed with the ADHD section of the Structural Clinical Interview for DSM-IV, Childhood Diagnosis (KID-SCID) to examine whether ADHD was present during primary school age to exclude false positives.

4.4.3 Sample size

Previous findings of Farrell et al., 2009 indicate an outcome difference between GST and TAU with an effect size of d around 2 (d refers to Cohen's d (Cohen, 1988)). However, some shrinkage of the effect might be expected when GST is provided by centres that were not involved in the development of GST. Also, although TAU was virtually ineffective in the abovementioned study, recent meta-analyses and RCTs of modern treatments for BPD suggest that these treatments can be effective (McMain et al., 2009; Perry et al., 1999). Hence, the current RCT is designed to have sufficient power to detect an effect size of $d = 0.5$.

With respect to the comparison between the two formats for GST, no large outcome differences are expected. Whereas small differences are unlikely to influence the preferences of patients and therapists, medium differences may. Therefore, sufficient power is needed to detect an effect size of $d = 0.5$ between the two formats of GST.

Over the course of three years a dropout of 20% is expected, with about 5% in the first year. Assuming the effects of GST and TAU become apparent after one year (Farrell et al., 2009), the study is powered taking into account a 5% attrition rate. Later attrition is partly compensated by including all randomized patients in analysis. Furthermore, since conservative effect sizes were chosen, the calculated initial sample size could be larger than needed and may therefore also compensate for attrition over 5%.

To test whether GST is superior to TAU and assuming centre as random effect and a centre by treatment interaction that corresponds to a typical intraclass correlation coefficient (ICC) of almost 0.05 in line with literature (Adams et al., 2004; Eldridge et al.,

2004), 236 patients (which implies 8 centres of 32 patients, 16 receiving TAU and 16 receiving GST) are needed to detect a difference of d = 0.5 with 90% power using a two-tailed significance level of $\alpha = 0.05$. For the comparison between GST-A and GST-B and again assuming a typical intraclass correlation of nearly 0.05, a sample size is needed of 202 patients (which implies 13 centres of 16 patients, 8 receiving GST-A and 8 receiving GST-B) to have 90% power to detect a difference of d = 0.5 at a two-tailed significance level of $\alpha = 0.05$. Taking into account the expected 5% attrition, one additional centre is needed. To be able to balance the orders in which GST formats are delivered to successive cohorts, an even number of sites is required. Including 14 centres gives 90% power to detect a difference of d = 0.40 between GST and TAU and a difference of d = 0.50 between GST-A and GST-B, both at a two-tailed significance level of $\alpha = 0.05$ (a detailed explanation of the sample size calculation is provided in Appendix 1).

4.4.4 Randomization

Patients will be randomized centrally in blocks of two per centre (GST versus TAU) using a computer-generated list by an independent central research assistant after baseline screening is complete and all in/exclusion criteria have been checked by this assistant. Each centre will have at least two cohorts of at least sixteen patients of which eight are randomized to GST and eight to TAU. In half of the centres the patients from the first cohort will be randomized to GST format A, which is an intensive group treatment or TAU. The patients from the second cohort will be randomized to GST format B, which combines individual and group treatments or TAU. In the other half of the centres the first cohort will be randomized to GST-B or TAU and the second to GST-A or TAU. The GST formats as well as the orders in which they are delivered to successive cohorts (first GST-A versus TAU, then GST-B versus TAU, and vice versa) will be balanced. Within countries that contribute an uneven number of sites, the numbers of orders of GST formats will have a difference of one.

4.4.5 Assessments

Prior to randomization, patients will be assessed at baseline. Baseline assessments will be spread over a period of approximately three months on average. Once a cohort is nearly complete, baseline assessments can be speeded up so they are completed within a minimum time period of one month. Inversely, when inclusion is too slow, the maximum time period patients are required to wait before treatment will start is limited to one year. When baseline assessments are complete for all patients, patients will be randomized and treatment will start. Subsequently, patients will be re-assessed approximately

every six months over the course of two years. Since the treatment duration for GST is also two years, the fourth assessment after baseline will coincide with GST treatment ending. Follow-up assessments will take place one year later. For the cost interview, a recall period of a year is considered too long. Therefore, an additional assessment will take place midway during the follow-up time period (i.e., two-and-a-half years after start of treatment) during which only the cost interview will be performed.

All assessments will be performed by local research assistants in the centres, except for SCID-interviews which will be done by trained interviewers blind for condition. Assessments include PC-based self-report questionnaires and interviews. An overview of the instruments used per assessment is provided in Table 4.1.

Table 4.1: Instruments used per assessment

	Screening	Baseline[1]	6 months	12 months	18 months	24 months	30 months	36 months
SCID I	•							•
SCID II	•							•
WHO ADHD screener	•							
ITEC	•							
BPDSI-IV		••	•	•	•	•		•
GAF		••	•	•	•	•		•
SOFAS		••	•	•	•	•		•
BPD checklist		•	•	•	•	•		•
WSAS		•	•	•	•	•		•
BSI		•	•	•	•	•		•
YSQ-short form		•	•	•	•	•		•
SMI		•	•	•	•	•		•
RSQ		•	•	•	•	•		•
ECNI		•	•	•	•	•		•
EuroQol-5D		•	•	•	•	•		•
WHOQOL-short version		•	•	•	•	•		•
Cost interview		•	•	•	•	•	•	•

Abbreviations: SCID I = Structured Clinical Interview for DSM-IV Axis-I Disorders, SCID-II = Structured Clinical Interview for DSM-IV Axis-II Disorders, WHO ADHD = World Health Organization Attention Deficit Hyperactivity Disorder, ITEC = Interview for Traumatic Events in Childhood, BPDSI-IV = Borderline Personality Disorder Severity Index version IV, GAF = Global Assessment of Functioning, SOFAS = Social and Occupational Functioning Assessment Scale, BPD = Borderline Personality Disorder, WSAS = Work and Social Adjustment Scale, BSI = Brief Symptom Inventory, YSQ = Young Schema Questionnaire, SMI = Schema Mode Inventory, RSQ = Relationships Scales Questionnaire, ECNI = Emotional Core Needs Inventory, EuroQol-5D = European Quality of Life questionnaire-5 dimensions, WHOQOL = World Health Organization Quality of Life questionnaire.

[1] Baseline consists of five assessments and BPDSI, GAF and SOFAS are assessed twice at baseline.

At baseline, the Interview for Traumatic Events in Childhood (ITEC) will be conducted. This is a retrospective, semi-structured interview for childhood maltreatment including sexual, physical and emotional abuse and physical and emotional neglect (Lobbestael et al., 2009). At follow-up, patients are assessed for recovery from BPD as well as the most common comorbidities by trained interviewers using the following sections of SCID-I for DSM-IV-Tr: affective disorders (including bipolar disorder I and II), anxiety disorders, eating disorders and substance disorders; and the following sections of SCID-II for DSM-IV-Tr: avoidant, dependent, obsessive-compulsive, paranoid, schizotypal, schizoid, histrionic, narcissistic, antisocial and borderline PD. Since only current diagnoses need to be considered for the assessment of recovery from BPD and comorbidity, the recall period for these shortened SCIDs is limited to six months. Only when a patient will otherwise become a study dropout (e.g., due to unwillingness to come to the centre for assessments), they can fill in questionnaires at home. Interviews are then conducted by telephone. By doing so, the occurrence of missing data will be reduced. Due to the nature of the study, blinding of participants is not possible. Except for the cost interview, which contains specific questions on which treatments patients have received, assessments will be performed by blinded local research assistants. The cost-interview, containing questions on health care utilization that would unblind the condition will be done by a nonblinded research assistant. This assistant will also monitor treatments provided in TAU. Furthermore, this nonblinded research assistant will collect treatment session recordings that are needed for supervision and validation of treatment adherence. After checking the quality of the recording, the sample of recordings that is needed will be stored.

To maximize adherence to the study protocol, a manual has been created for all local research assistants. A central research assistant is appointed to whom local research assistants can address questions concerning any logistical issues. This individual will also perform checks and provides guidance and direction when needed. Checks include study protocol adherence, in- and exclusion criteria of candidate participants, and keeping track of the scheduling and advancement of assessments, data and the collection of audio and video recordings. The central research assistant will furthermore train local research assistants, distribute the instruments used for assessments, provide updates of the manual, explain data format requirements, and prepare the online data collection environment. In sum, the central research assistant will safeguard the validity of assessments.

4.4.6 Treatments

4.4.6.1 Group schema therapy

Schemas refer to the knowledge representations people have about themselves, others and the world and which are formed during childhood by the way basic needs are met. When a child has to cope with his or her basic needs not being met, a variety of dysfunctional schemas, and/or coping styles, may develop. Schema modes refer to the emotional states between which BPD patients can rapidly switch when triggered by events that are related to the unmet needs during childhood. Schema modes represent sets of schemas and/or coping styles that are typically expressed in such emotional states. The following schema modes are characteristic for BPD: the vulnerable (abandoned/abused) child, angry/impulsive child, punitive parent, detached protector (or any other protective mode), healthy adult and happy or contented child. The first four of these modes are maladaptive and strongly present in BPD patients. The latter two are functional modes and weak in BPD patients. Schema therapy aims to reduce maladaptive modes and develop and strengthen functional modes (Kellogg & Young, 2006). This schema mode model guides therapy as each mode requires a different treatment strategy. The strategies comprise specific experiential, cognitive and behavioural techniques (Jacob & Arntz, 2003). When offered in a group format, several factors may amplify and speed up recovery: peer support, a sense of belonging and understanding, opportunities for vicarious learning and real-life practice of healthy behaviour (Farrell et al., 2009).

In format A (GST-A), two-year GST consists of 124 groups sessions with a duration of 90 minutes. In the first year group sessions take place twice a week and in the second year the frequency gradually decreases. In addition, in GST-A a total of up to 18 individual sessions can be used at the patients discretion or in times of crisis. Two individual sessions take place before group sessions commence. At this time patients get acquainted with their group schema therapists, schema therapy and the schema mode model are explained, the schema modes a patient has and their relationship to one another are identified and a treatment plan is drawn up.

In format B (GST-B), two-year GST involves a combination of group and individual sessions. In the first year, there are weekly group sessions of 90 minutes and individual sessions of 50 minutes and in the second year the frequency gradually decreases. In total, patients in this condition receive 74 group sessions and 62 individual sessions. In the first two individual sessions patients get acquainted with their individual and group therapists, schema therapy and the schema mode model are explained, the schema modes a patient has and their relationship to one another are identified and a treatment plan is drawn up. Group and individual ST therapists meet regularly to coordinate treatments.

4.4.6.2 Treatment as usual

Following usual procedures, the intake staff at each centre determine the optimal treatment offered to each patient in the TAU condition. Except for (G)ST, the intake staff are allowed a choice from the whole array of possible treatments for BPD with no restriction; whether intensive, individual, group, inpatient, outpatient or day treatment. The TAU condition is thus representative of optimal current practice and will be carefully monitored for all patients. Since there is no protocol for TAU in this RCT and specific treatment is decided on by experts in the community treatment centre, TAU is equivalent to 'community treatment by experts'.

In some centres, the treatment that is usually offered to BPD patients is Dialectical Behavioural Therapy (DBT). This is an empirically validated treatment for BPD (Lynch et al., 2007; McMain et al., 2009). If the number of centres that offer DBT as the usual treatment is sufficient, this will provide an opportunity to compare the clinical effectiveness and cost-effectiveness of GST with DBT.

4.4.6.3 Therapists, training and supervision

GST sessions are run by two schema therapists, of whom at least one is a senior schema-therapist. Stand-in schema therapists replace the regular therapists when they are absent (e.g., due to illness, holidays or pregnancy). The senior therapist's role, in addition to being a group schema therapist, is to act as a local supervisor for other schema therapists. Individual schema therapists can also be group schema therapists for GST. Group schema therapists receive a training of six days in GST (Farrell et al., 2012). Candidate ST therapists (individual and group) who are not yet trained in ST first receive a 3-4 day training in individual ST for BPD (Arntz et al., 2009). During the study intensive supervision meetings are held twice a year in the first year and once or twice a year in the second year. In addition, weekly team supervision is provided locally and central supervision by the developers of GST is provided through teleconferencing and viewing of video-recordings weekly in the first months, then biweekly and monthly after about 6 months. Initially, a computer program was acquired that was especially designed to enable secure online sharing of video-recordings of medical procedures through encrypted streaming. Unfortunately, this program has become unavailable. The encrypted recordings will now be uploaded on a central server so that supervisors can download and watch them.

4.5 Evaluation of clinical effectiveness

In this section, the primary and secondary clinical outcome measures that will be used to evaluate the clinical effectiveness of the treatments are described as well as the strategy used for analysis of the data. All instruments that will be used to investigate clinical effectiveness that were not yet available in the languages of all participating sites were translated. Translated versions were backtranslated and thoroughly checked for consistency with the original version. Any inconsistencies were addressed in consensus meetings and adjusted accordingly. Instruments are implemented in an online data collection environment.

4.5.1 Clinical outcome measures

4.5.1.1 Borderline Personality Disorder Severity Index version IV (BPDSI-IV)

The primary outcome measure is the severity of BPD, expressed as a score between 0 and 90 as measured with the Borderline Personality Disorder Severity Index (BPDSI), version IV. The BPDSI-IV is a semi-structured interview containing 70 items based on the nine BPD dimensions described in DSM-IV. This is a reliable and valid instrument, suitable for use as an outcome measure (Arntz et al., 2003; Giesen-Bloo et al., 2010). A cut-off score of 15 between patients and controls has been previously established (Arntz et al., 2003; Giesen-Bloo et al., 2010). Therefore, a score below 15 measured two years after randomization or earlier and maintained until follow-up can be used as a criterion for recovery. The scores on subscales of the BPDSI-IV provide information on the severity of each of the nine dimensions of BPD. The recall period for the BPDSI-IV is three months.

4.5.1.2 BPD checklist

The BPD checklist is a self-report instrument that measures the burden of BPD manifestations as experienced by patients. Since the BPD checklist measures changes in subjective burden, it is complementary to the BPDSI-IV that measures changes in symptomatology objectively. It consists of 47 items based on the nine dimensions of BPD in DSM-IV and answers are scored on a five point Likert scale. Suitability for use as a treatment outcome measure has been established (Giesen-Bloo et al., 2006). The recall period for the BPD checklist is one month.

4.5.1.3 Brief Symptom Inventory (BSI)

The BSI is a self-report instrument used as an inventory of general psychiatric symptoms present at the time of assessment and is a short alternative to the SCL-90-R, from which it was developed (Derogatis & Melisaratos, 1983). It contains 53 items to inventory the following nine types of primary symptom dimensions: somatic, cognitive, interpersonal sensitivity, depressive mood, anxiety, hostility, phobia, paranoia and psychoticism. Answers are scored on a 5-point Likert Scale. Cronbach's α is .96 for the instrument in total and ranges between .71 and .87 for its subscales (de Beurs & Zitman, 2006). According to Bland & Altman, 1997, Cronbach's α values of 0.9 or higher indicate an internal consistency that is appropriate for clinical applications, and values of 0.7 to 0.8 are satisfactory for comparing groups. In addition, the BSI has good discriminant validity (de Beurs & Zitman, 2006).

4.5.1.4 Happiness item

The happiness item is a single question on general happiness in the months prior to the assessment and is scored on a seven point Likert scale (Veenhoven, 2008). This scale consists of the following verbal descriptions of different states of happiness: (1) completely unhappy, (2) very unhappy, (3) fairly unhappy, (4) neither happy nor unhappy, (5) fairly happy, (6) very happy, (7) completely happy. Norms for all participating countries are available (Veenhoven, 2008). For a single happiness item high test-retest reliability ($r = 0.86$) and good concurrent, convergent, and divergent validity have been reported (Abdel-Khalek, 2006). The happiness item has excellent sensitivity to change for patients with BPD who were treated with GST (Dickhaut & Arntz, 2014).

4.5.1.5 Schema Mode Inventory (SMI)

The SMI is a self-report instrument that consists of 143 items on 16 schema modes that are scored on a six point Likert scale. It measures the extent to which dysfunctional as well as functional schema modes are present at the time of assessment (Lobbestael et al., 2008). It is an adaptation of the original SMI containing 270 items (Young et al., 2007) and short SMI containing 118 items (Lobbestael et al., 2010). This instrument is only used for patients in the GST condition. Its subscales have satisfactory to high internal consistency (Cronbach's α ranges from .79 to .96) (Lobbestael et al., 2010). The SMI is a useful instrument for assessing modes in ST (Lobbestael et al., 2010).

4.5.1.6 Young Schema Questionnaire - short form (YSQ)

The YSQ is a self-report instrument containing 75 items that are scored on a six point Likert scale (Young, 1998). It is used to measure the presence or absence of 16 core maladaptive schemas at the time of assessment. The YSQ has adequate temporal as well as rank-order stability and an analysis of its discriminant power in clinical versus non-clinical samples revealed it is highly sensitive in predicting the presence or absence of psychopathology (Rijkeboer et al., 2005). Internal consistency is high for the overall scale (Cronbach's α ranges from .94 to .96) and satisfactory to high for its subscales (Cronbach's α ranges from .72 to .94) (Baranoff et al., 2006).

4.5.1.7 Global Assessment of Functioning (GAF) and Social and Occupational Functioning Assessment Scale (SOFAS)

Based on axis V of DSM-IV, the GAF and SOFAS are 100-point scales used to assess general and social/occupational functioning, respectively. A short semi-structured interview serves to elicit the information needed for scoring. The recall period for both instruments is one month. The GAF is a valid scale of global psychopathology and the SOFAS is a valid measure of social, occupational and interpersonal functioning (Hilsenroth et al., 2000). Both instruments have excellent interrater reliability (intraclass correlation coefficients $>$ .74) (Hilsenroth et al., 2000).

4.5.1.8 Work and Social Adjustment Scale (WSAS)

The WSAS is a self-report instrument that consists of 5 items that are scored on a scale ranging from 0 to 8. It is used to assess functional impairment at the time of assessment in the domains of work, household, social leisure, private leisure and family and relationships. The WSAS' reliability, validity and sensitivity to change have been firmly established in samples of patients with different clinical disorders (Mundt et al., 2002; Mataix-Cols et al., 2005; Jansson-Frojmark, 2014).

4.5.1.9 Relationship Scales Questionnaire (RSQ)

The RSQ was designed as a continuous measure of adult attachment and consists of 30 short statements about romantic relationships (Griffin & Bartholomew, 1994; Scharfe & Bartholomew, 1994). After being instructed to think about such relationships in their past and present lives patients rate the extent that these statements resemble their own feelings and experiences at the time of assessment on a five point Likert scale. Scores are calculated for the following attachment patterns: secure, dismissive, fearful and preoccupied.

4.5.1.10 Emotional Core Needs Inventory (ECNI)

The ECNI is a list of 88 statements used to measure the extent to which basic needs are being met in important relationships at the time of assessment (Perris et al., 2008). Each statement is rated on a scale ranging from 1 to 6. A higher rating corresponds to better need-fulfillment. Assessing the extent to which patient's needs are met by others is important given that maladaptive schemas are hypothesised to result from unmet core emotional needs. A central aim of ST is to bring about changes that lead to better need-fulfillment of patients (Lockwood & Perris, 2012).

4.5.1.11 World Health Organization Quality of Life questionnaire (WHOQOL-short)

The WHOQOL-short is a self-report instrument for assessing quality of life in the two weeks prior to assessment. It is a short version (35 items) of the WHOQOL and focuses on the domains of physical health, psychological health, social relationships, environment, positive feelings, negative feelings and self-esteem. The WHOQOL-short is a reliable and valid instrument ("Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group", 1998).

4.5.2 Analyses

All available data on clinical outcomes will be analysed according to the intention-to-treat principle. Outcome measures will be analysed with mixed regression, also known as multilevel or hierarchical regression, with centre as a random effect, allowing centre by treatment interaction, and including patient-level and treatment indicator covariates, as well as time and treatment by time effects. Baseline covariates will be used to adjust for potential differences at baseline and to reduce standard errors. For categorical outcome variables, counts and in case of nonnormal residuals, appropriate forms of mixed regression will be chosen (binomial, Poisson, gamma, etc.).

4.6 Economic evaluation

The following section describes how costs and utilities will be measured for the economic evaluation as well as the planned cost-effectiveness and cost-utility analyses. The cost interview has been translated into the languages of all participating sites. The original Dutch version was first translated to English and subsequently this English version

was then translated into Greek and German. The translated versions have been back-translated and/or thoroughly checked for consistency with the original version. Any inconsistencies were addressed in consensus meetings and adjusted accordingly. The cost interview will be implemented in an online data collection environment.

4.6.1 Costs measurement

Costs will be measured from a societal perspective using a retrospective cost interview especially designed for BPD patients. Relevant costs to be identified include healthcare, patient and family costs, and costs outside the health care sector. Healthcare costs include visits to general practitioners, hospitals, psychiatrists and psychologists, crisis centres, use of medication, social work, formal care, and alternative treatments. Patient and family costs include travelling costs, informal care (care provided by family, friends or neighbours of the patient) and out of pocket costs (alcohol, drugs, smoking and self-reported other costs). Costs in other sectors include productivity losses from unpaid work (voluntary work and study) and paid work. The cost interview will be conducted by trained local research assistants, who will ask questions on the consumption of different resources and assess volumes of resource use. When applicable, calculations and descriptive content will be noted. For the cost interview the recall period is 6 months.

4.6.2 Utility measures

4.6.2.1 EuroQol-5D-3 L (EQ-5D-3 L)

The EQ-5D-3 L is a generic, self-assessment instrument used for measuring health-related quality of life at the time of assessment (Brooks, 1996). It consists of five questions, each related to a specific dimension of health status: mobility, self-care, usual activity, pain/discomfort and anxiety/depression. The EQ-5D has been translated in the languages of all participating sites (*www.euroqol.org*, 2014). The descriptive profiles, generated by the EQ-5D-3 L are valued using social tariffs for the EuroQol to generate utilities, which reflect a population's preference for a particular health profile.

In base-case analysis, country-specific tariffs will be used for valuation when available. For the Netherlands, Germany and the UK national tariffs are available, whereas for Greece and Australia they are not (*www.euroqol.org*, 2014). For these countries, proxy tariffs will be used as an alternative (e.g., tariffs for Europe and New-Zealand, respectively). Sensitivity analyses will be performed that make use of country-specific tariffs for the valuation of health profiles of all patients.

Utilities will be used to calculate Quality Adjusted Life Years (QALYs) by multiplying change in utility between assessments by the duration of the time period between

assessments. Through the use of statistical regression, potential baseline differences in QALYs can be adjusted for (Manca et al., 2005).

In addition, the EuroQol thermometer will be scored in a range between 0 and 100 to provide a single index measure for a patient's health status (Brooks, 1996).

4.6.3 Analyses

The economic evaluation will be comprised of both a cost-effectiveness analysis (CEA) and cost-utility analysis (CUA) and will be performed from a societal perspective. All available data on costs and outcomes will be analysed according to the intention-to-treat principle. Data gathered with the cost interview will first be checked for adherence to questionnaire routing, illogical answers, unrealistic cost estimates and outliers. For any problems in the data that cannot be solved through logic, decision rules will be established. A decision rule could apply a specific limit to volumes of costs, e.g., the maximal hours patients spend per day on caring for their children can be limited to eight for patients who report doing this for 24 hours per day. For intermittent missing values at the item level of the cost interview the mean values of previous and subsequent assessments will be imputed to allow calculation of total costs. Missing values on total scores or other outcomes will not be imputed. Unrealistic or extreme values will be investigated per case and corrected when appropriate. Volumes of resource use as measured by the cost interview will be multiplied by their corresponding unit costs and then summed to provide an overall total cost. Unit costs will be based on standard unit prices for each country (e.g., in the Netherlands: Hakkaart-van Roijen et al., 2010 for cost prices of healthcare services) when available and on (averaged) tariffs otherwise. Cost prices will be expressed in Euros for the same base year and indexed using consumer price indices when required. Cost prices expressed in currencies other than Euros will be converted using purchasing power parities. In addition, to account for the three year time horizon of this RCT, cost prices will be discounted according to the guidelines. When neither standard unit prices nor tariffs are available for specific resource items in specific countries, alternative strategies for the valuation of resource use will be considered. When for a subset of resource items the unit costs are known in all countries and for all other resource items there is a unit cost available in at least one country, the unit cost of a resource item that is not available in one or more countries can be estimated through a procedure called market-basket based imputation (Schulman et al., 1998). In base-case analysis, country-specific unit costs will be used to value resource use for each country. Sensitivity of the results to differences in unit prices between countries will be analysed by valuing resource use in all countries using unit prices of a single country. Productivity losses will be valued by the human capital approach through multiplication of the total number of hours lost

with the national average hour wage. Shadow prices will be used to value informal care and lost productivity in study and voluntary work. The primary clinical outcome for the CEA will be the BPD Severity Index (BPDSI) score and for the CUA primary utility scores will be derived from the EuroQol-5D. Cost-effectiveness and cost-utility data will be analysed with multilevel modelling techniques. The net monetary benefit will be used to express cost-effectiveness, and results will be expressed in cost-effectiveness acceptability curves. Net monetary benefit (NMB) will be calculated for a range of values for the amount decision makers are willing to pay (WTP) for an additional unit of effect, following: $NMB = \Delta E * WTP - \Delta C$, where ΔE is the difference in effects and ΔC is the difference in costs (Hoch et al., 2002). To accommodate the skewness that is typically observed in cost data, costs (or NMB) can be assumed to follow a gamma or log-normal distribution. Baseline covariates will be used to adjust for potential differences at baseline and to reduce standard errors. Results will be expressed in cost-effectiveness acceptability curves that display the probability, based on the available evidence, that GST can be considered cost-effective for a range of WTP values. For these types of analyses, a Bayesian approach is the most natural, since it allows direct and intuitive statements to be made regarding the probability that a treatment is cost-effective, based on the available evidence (O'Hagan & Luce, 2003). Furthermore, Bayesian methods allow flexible joint modelling of costs and effects, thereby facilitating sensitivity analyses regarding different methodological approaches for specification and parameterization of the model. Sensitivity analyses will be performed to address uncertainty regarding methodology, model specification and prior distributions.

4.7 Additional substudies

Complementary to effectiveness and cost-effectiveness evaluations, a series of additional investigations will be performed that consist of an assessment of the integrity of GST, an investigation into the opinions of primary stakeholders and additional studies that examine variables that might mediate the change process in GST.

4.7.1 Treatment integrity

Adherence to GST treatment protocol will be assessed by trained independent judges who rate a random selection of video-recordings of group-ST using a newly developed instrument. For individual ST, a random selection of video-recordings will be rated using existing instruments (Bamelis et al., 2014; Giesen-Bloo et al., 2006) by trained raters. As sampling recordings of TAU will be impossible in many sites, due to ethical and lo-

gistic reasons (e.g., TAU in private practice or TAU in mixed groups including patients not participating), the differentiation between ST and TAU will be assessed by having patients fill out a checklist with techniques that are typical for ST and non-typical for ST.

4.7.2 Patient and therapist perspectives

This study involves the acquisition of qualitative data in the form of patients' and therapists' opinions about treatment and the preferred format for GST. The opinions of these major stakeholders will be elicited through in-depth interviews and/or focus groups, allowing them to share their view on treatment and the preferred format for delivery of GST. Topics include which aspects of the GST protocol are perceived to be beneficial or not, application of specific ST techniques, therapeutic relationships and therapist training and supervision. Saturation is expected to be reached after having interviewed 12 GST patients in each participating country and 12-15 therapists, after which assessment will be discontinued. The centres that participate in this aspect of the study are from the Netherlands, Germany and Australia. Centres in other countries might decide to participate later. Patients will be sampled from both GST formats and from 3 phases of treatment: first year, second year, and after treatment completion. All interviews are recorded. After a full transcription of the recordings is made a summary is made. This summary is then reviewed by patients and therapists as a check on its veracity. If needed, the summary will be corrected to be sure that the verbatim transcripts express patient and therapist opinions correctly. Verbatim transcripts of interviews with patients and therapists will be analysed for their content using specialized software. Important and/or recurring themes will be categorized, interpreted and reported.

4.7.3 Studies on variables affecting treatment outcomes

The final area for additional investigation is the extent that patient variables affect outcomes such as dropout rates and patient improvement. This includes genetic polymorphisms, dissociation, comorbidity, individual patient trajectories, change processes and the therapeutic relationship, changes in neural correlates of emotional sensitivity, regulation and impulsivity during treatment, changes in threat bias during treatment, changes in needs during treatment, somatic symptoms and somatization, the therapist's voice and use of recordings for secure attachment, the effects of training, early self-understanding as a predictor for outcome, the effect of treatment on comorbidity and changes in attachment representation.

4.8 Discussion

In this article the design is described for an international, multicentre RCT on GST that includes an evaluation of the clinical effectiveness, a full economic evaluation as well as a series of additional investigations. In this RCT, GST (format A and B) will be compared against optimal TAU in terms of clinical effectiveness and cost-effectiveness. Such a design follows the 'gold standard' in cost-effectiveness research (Gold et al., 1996) and allows us to investigate whether GST excels current practice, which consists of the existing optimal treatments that are usually provided to patients with BPD. TAU consists of a variety of different treatments due to the fact clinical practice varies between centres as well as between countries. Since TAU is tailored to the individual needs of each patient, it can be considered representative of optimal current practice. The multicentre and international design of this RCT specifically intends to capture the variation in clinical practice between participating centres and between countries, respectively. Because the resulting amalgam of treatments in TAU reflects current practice, external validity is increased. If the RCT was designed to include a fixed treatment instead of variable TAU as a comparison to GST, then it would be less informative in regard to whether GST excels current practice and whether its further implementation is supported. Furthermore, if the RCT was designed as a head-to-head comparison between GST and another experimental treatment without the inclusion of TAU, then interpretation of its results could be hindered. This is because it is not clear how experimental treatments compare to TAU. For instance, experimental treatments might do worse than or be equivalent to TAU. For these reasons, TAU is considered to be an appropriate comparator to GST at this stage of research. However, with this TAU there is little control over the specific issues that are addressed in therapy, the amount of attention a patient receives and the frequency of therapeutic contacts. It is therefore less rigorously defined than the experimental condition. In addition, therapists providing TAU may not receive the intensive supervision that GST therapists receive and the treatment fidelity of the components of TAU is not monitored. These issues could be a potential threat to internal validity (Mohr et al., 2009). Notwithstanding, TAU will be delivered by skillful therapists with extensive experience in the treatment of BPD and its contents are monitored by administering a questionnaire on the specific treatments that each patient receives.

The fact that this RCT will take place with multiple participating centres and in an international context has specific advantages and disadvantages. Several advantages of international clinical trials over single-country trials have been formulated (Pang, 2002), which also hold in the case of multicentre trials versus single centre trials. First, in multicentre or international RCTs, through parallel recruitment of patients at the different sites, it takes less time to include a sufficient number of patients in comparison to single

site studies. Second, the representativeness of the study population is enhanced by capturing more of the variability in patient characteristics, clinical practice patterns and/or health care systems. Third, the collection of data at multiple sites enables the researcher to inform decision makers in all of the participating sites.

Paradoxically, whereas the inclusion of more variability enhances the representativeness of the study population, this same variability makes it difficult to apply the results to any one centre or country in particular. In other words, studies designed to include multiple participating centres and/or countries raise issues concerning their generalizability (Manca & Willan, 2006; Sculpher et al., 2004). At the patient level, variation between sites exists in terms of demography and epidemiology. At the level of treatment centres and clinicians, differences may exist in patient management. Differences between healthcare systems and other socioeconomic factors may influence healthcare delivery and the allocation of scarce resources to healthcare, respectively. Inversely, some degree of similarity within healthcare systems, treatment centres or patients may also be expected. A method that can accommodate the hierarchical structure of such data is multilevel modelling, which has been proposed as an appropriate analytic strategy for cost-effectiveness data from multinational RCTs (Drummond et al., 2005; Manca & Willan, 2006; Manca et al., 2010; Sculpher et al., 2004). It allows variation to be estimated within and between the different levels. Moreover, these estimates can be used to calculate centre-specific estimates of cost-effectiveness (Sculpher et al., 2004), which can be used to determine the extent to which the results from this RCT are generalizable.

Since this RCT involves multiple participating centres in different countries, organizational and logistical challenges potentially threaten its quality. Handling these challenges is a labour-intensive process that requires thoughtful planning, a clear protocol, continuous monitoring of protocol adherence, and well-defined communicational lines. The hub in the logistical infrastructure is a central research assistant who will perform checks and steering concerning study protocol adherence and therefore plays an important role in ensuring the validity of the assessments. Organizational issues may arise when, for example, therapists, coordinators or research assistants retreat from the study and are replaced, when recruitment rates are slower than was foreseen or video-facilities for treatment supervision are missing. The appropriate handling of these issues requires timely noticing of their occurrence, which will be facilitated by regular internet conferences involving the principal investigators and local coordinators.

In this RCT, both primary and several secondary clinical outcome measures will be assessed through interviews. Therefore, it is necessary to control interviewer bias by having blinded interviewers perform the assessments. For interviews containing specific questions about which treatment the patient receives (the cost interview and the monitoring of treatments provided in TAU), blinding of interviewers will obviously not be

possible. These interviews will be performed by non-blinded research assistants. All other interviews and questionnaires will be performed and administered by blinded research assistants.

When conducting an economic evaluation in the field of mental health, it can be a challenge to provide a comprehensive account of all the costs and consequences that are associated with the treatments being compared and which are relevant under the target perspective (Knapp, 1999). Since relevant costs for BPD include health care costs, patient and family costs and costs in other sectors, a societal perspective will be taken. This also prevents cost shifting to be interpreted as increases or decreases in costs. The time horizon of three years covers both the duration of the GST treatment as well as a one year follow-up time period. This enables an investigation into the stability of treatment outcome over time. In addition, relevant costs that are incurred once the GST treatment has ended, which could be a consequence of treatment outcome, are thus included. To gather data on the societal costs and consequences that are associated with BPD and the BPD treatments being compared in this RCT, resource use in a wide range of health care facilities is taken into account, whether inpatient or outpatient, including various health services specialized for mental health (e.g., contacts with a psychologist or psychiatrist) as well as more general health care services (e.g., contacts with a GP or general hospital). Furthermore, costs due to productivity losses and informal care will be measured to take into account the fact that costs and consequences have an impact on society as well as family and friends, respectively. Lastly, by taking into account various categories of out-of-pocket expenses that are typically associated with BPD (e.g., alcohol, tobacco and drug use, impulsive buying, binge eating) an attempt will be made to measure all relevant costs that are specifically associated with BPD.

Despite the extensive effort that is put into obtaining a complete picture of the costs and consequences that are associated with BPD and the BPD treatments being compared in this RCT, it remains unfeasible to include some particular aspects. For instance, to date no instruments exist to measure the high burden that BPD patients can impose upon colleagues and organizations due to suboptimal functioning at work (Bamelis, 2013). Also, leisure time is relevant to patients with BPD, but it can be difficult to measure and value. Therefore, these aspects are not taken into account as costs or consequences in the economic evaluation.

Another degree of complexity is added to economic evaluations in the field of mental health, as opposed to somatic disorders, due to the fact that once costs and consequences are measured, their interpretation can be difficult. No consensus exists on the extent to which particular costs need to be included or not and how they need to be valued (McCrone, 2011). For example, although informal care provided by family and friends is very relevant to patients with BPD, it can be difficult to know which amount of care is

specifically due to mental health problems and which amount they would have received anyway. Similarly, productivity losses can be the result of being under treatment, yet patients with mental health problems are also less likely to be employed or could have already lost their job before having their diagnosis (Evers et al., 1997). Furthermore, although the volumes of production losses in paid work, voluntary work, study, and household activities as well as contacts with the GP and medication use are explicitly measured separately for being BPD-related or not, this division cannot always be reliably made. For example, a leg injury by itself may seem unrelated to BPD at first sight, but less so when it is the result of a suicide attempt. In such cases, the analyst is guided by the information provided, while at the same time being aware that this information may or may not be complete. In cases where there is sufficient information to attribute a somatic complaint to an underlying mental health complaint, patient answers are overruled. Reliability can also be an issue when rather high out-of-pocket expenses are reported (Bamelis, 2013).

In addition to the evaluation of clinical effectiveness and the economic evaluation, a series of additional investigations will be performed in this RCT that consist of an assessment of the integrity of GST, in terms of adherence to the GST protocol, an investigation into the opinions of major stakeholders and analyses of variables that might mediate treatment response. The qualitative data on the experiences of patients and therapists are considered as complementary to the quantitative methods that will be employed in this study. By interviewing patients and therapists, potentially important, yet unanticipated, issues may be detected. Furthermore, this type of data collection can give valuable insight into the contextual factors that play a role in the effectiveness of GST and its implementation.

4.9 Conclusion

GST holds much promise as a treatment for BPD. However, since only two small studies have tested GST important questions remain to be answered before its further implementation is supported. The current international, multicentre RCT is designed to reveal how GST, when delivered as a complete and stand-alone treatment by therapists who were not involved in its development, compares to up-to-date TAU. Concurrently, this RCT aims to investigate the optimal format for the delivery of GST; whether consisting almost exclusively of group sessions or as a combination of individual and group sessions.

In addition to an investigation of clinical effectiveness, this international multicentre RCT will involve an economic evaluation to investigate how GST compares to TAU, and how both formats for GST compare in both clinical and economic terms. Furthermore,

a series of additional investigations will be performed to shed light on the qualitative aspects of GST and on variables that influence treatment outcome. In sum, this RCT contributes to an evidence-based understanding that will inform decisions regarding which treatment to offer to patients with BPD, both from a clinical and societal perspective.

4.10 Acknowledgements

The authors would like to thank all patients, research assistants and therapists for their involvement in the study. We are thankful to Michelle van de Glind-Houben for central research assistance and Rosanne Janssen for developing an online data collection system. AA obtained funding for the Dutch sites of this study from the Netherlands Organization for Health Research and Development (ZonMW; 80-82310-97-12142) and Netherlands Foundation for Mental Health (2008 6350), GJ obtained funding from Else Kröner-Fresenius-Stiftung for the German sites, and CL obtained funding from Australian Rotary Health for the Australian site. The Greek site is supported by the Greek Society of Schema Therapy and the 1st Department of Psychiatry of the Medical School of the University of Athens. The site in London, UK, is supported by South London and Maudsley NHS Foundation Trust and Research Center Experimental Psychopathology (EPP), Maastricht University. The site in Bradford, UK, is supported by Bradford District Care Trust. The site in Indianapolis, USA, is supported by Midtown Mental Health Center and Research Center EPP, Maastricht University. Funding bodies played no role in the design of the study, in the collection, analysis and interpretation of data, in the writing of the manuscript and in the decision to submit the manuscript for publication.

4.11 References

Abdel-Khalek, A. M. (2006). Measuring happiness with a single-item scale. *Social Behavior and Personality: an international journal*, *34*(2), 139-150.

Adams, G., Gulliford, M. C., Ukoumunne, O. C., Eldridge, S., Chinn, S., & Campbell, M. J. (2004). Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *Journal of clinical epidemiology*, *57*(8), 785-94.

American Psychiatric Association: Practice guideline for the treatment of patients with borderline personality disorder. (2001). *Am J Psychiatry*, *158*(10 Suppl), 1-52.

Arntz, A., van den Hoorn, M., Cornelis, J., Verheul, R., van den Bosch, W. M., & de Bie, A. J. (2003). Reliability and validity of the borderline personality disorder severity index. *J Pers Disord*, *17*(1), 45-59.

Arntz, A., van Genderen, H., & Drost, J. (2009). *Schema therapy for borderline personality disorder*. Wiley.

Bamelis, L. L. (2013). *Considering care; a clinical and economic evaluation of schema therapy for personality disorders* (PhD thesis). Maastricht University.

Bamelis, L. L., Evers, S. M., Spinhoven, P., & Arntz, A. (2014). Results of a multicenter randomized controlled trial of the clinical effectiveness of schema therapy for personality disorders. *Am J Psychiatry*, *171*(3), 305-22.

Baranoff, J., Oei, T. P., Cho, S. H., & Kwon, S. M. (2006). Factor structure and internal consistency of the Young schema questionnaire (short form) in Korean and Australian samples. *J Affect Disord*, *93*(1-3), 133-40.

Bauer, R., Doring, A., Schmidt, T., & Spiessl, H. (2012). "Mad or bad?": burden on caregivers of patients with personality disorders. *J Pers Disord*, *26*(6), 956-71.

Bender, D. S., Dolan, R. T., Skodol, A. E., Sanislow, C. A., Dyck, I. R., McGlashan, T. H., … Gunderson, J. G. (2001). Treatment utilization by patients with personality disorders. *American Journal of psychiatry*, *158*(2), 295-302.

Bland, J. M., & Altman, D. G. (1997). Cronbach's alpha. *BMJ*, *314*(7080), 572.

Brooks, R. (1996). EuroQol: the current state of play. *Health Policy*, *37*(1), 53-72.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. L. Erlbaum Associates.

Coid, J., Yang, M., Bebbington, P., Moran, P., Brugha, T., Jenkins, R., … Ullrich, S. (2009). Borderline personality disorder: health service use and social functioning among a national household population. *Psychol Med*, *39*(10), 1721-31.

de Beurs, E., & Zitman, F. (2006). The brief symptom inventory (BSI): Betrouwbaarheid en validiteit van een handzaam alternatief voor de scl-90. *Maandblad Geestelijke Volksgezondheid*, *61*(2), 120-137.

Derogatis, L. R., & Melisaratos, N. (1983). The brief symptom inventory: an introductory report. *Psychol Med*, *13*(3), 595-605.

Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. (1998). *Psychol Med*, *28*(3), 551-8.

Dickhaut, V., & Arntz, A. (2014). Combined group and individual schema therapy for borderline personality disorder: A pilot study. *J Behav Ther Exp Psychiatry*, *45*(2), 242-251.

Drost, R. M. W. A., Paulus, A. T. G., Ruwaard, D., & Evers, S. M. A. A. (2013). Intersectoral costs and benefits of mental health prevention: towards a new classification scheme. *The Journal of Mental Health Policy and Economics*, *16*, 179-186.

Drummond, M., Manca, A., & Sculpher, M. (2005). Increasing the generalizability of economic evaluations: recommendations for the design, analysis, and reporting of studies. *Int J Technol Assess Health Care*, *21*(2), 165-71.

Dunne, E., & Rogers, B. (2013). "It's us that have to deal with it seven days a week": Carers and borderline personality disorder. *Community Ment Health J*.

Eldridge, S. M., Ashby, D., Feder, G. S., Rudnicka, A. R., & Ukoumunne, O. C. (2004). Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. *Clinical trials*, *1*(1), 80-90.

Evers, S. M., Van Wijk, A. S., & Ament, A. J. (1997). Economic evaluation of mental health care interventions. a review. *Health Econ*, *6*(2), 161-77.

Farrell, J., Shaw, I., & Arntz, A. (2012). *Group schema therapy for borderline personality disorder: A step-by-step treatment manual with patient workbook*. Wiley.

Farrell, J., Shaw, I. A., & Webber, M. A. (2009). A schema-focused approach to group psychotherapy for outpatients with borderline personality disorder: A randomized controlled trial. *Journal of Behavior Therapy and Experimental Psychiatry*, *40*(2), 317-328.

Feenstra, D. J., Hutsebaut, J., Laurenssen, E. M., Verheul, R., Busschbach, J. J., & Soeteman, D. I. (2012). The burden of disease among adolescents with personality pathology: quality of life and costs. *J Pers Disord*, *26*(4), 593-604.

Feurino, L., & Silk, K. R. (2011). State of the art in the pharmacologic treatment of borderline personality disorder. *Curr Psychiatry Rep*, *13*(1), 69-75.

Giesen-Bloo, J., van Dyck, R., Spinhoven, P., van Tilburg, W., Dirksen, C., van Asselt, T., … Arntz, A. (2006). Outpatient psychotherapy for borderline personality disorder: randomized trial of schema-focused therapy vs transference-focused psychotherapy. *Arch Gen Psychiatry*, *63*(6), 649-58.

Giesen-Bloo, J., Wachters, L., Schouten, E., & Arntz, A. (2010). The Borderline Personality Disorder Severity Index-IV: Psychometric evaluation and dimensional structure. *Personality and Individual Differences*, *49*, 136-141.

Gold, M. R., Siegel, J. E., Russell, L. B., & Weinstein, M. C. (1996). *Cost-effectiveness in health and medicine*. New York: Oxford University Press.

Griffin, D., & Bartholomew, K. (1994). The metaphysics of measurement: The case of adult attachment. In K. Bartholomew & D. Perlman (Eds.), *Advances in personal relationships: Attachment processes in adult relationships* (Vol. 5). London: Jessica Kingsley.

Hakkaart-van Roijen, L., Tan, S. S., & Bouwmans-Frijters, C. A. M. (2010). *Handleiding voor kostenonderzoek: Methoden en referentieprijzen voor economische evaluaties in de gezondheidszorg, geactualiseerde versie 2010.*

Hilsenroth, M. J., Ackerman, S. J., Blagys, M. D., Baumann, B. D., Baity, M. R., Smith, S. R., … Holdwick, J., D. J. (2000). Reliability and validity of DSM-IV axis V. *Am J Psychiatry*, *157*(11), 1858-63.

Hoch, J. S., Briggs, A. H., & Willan, A. R. (2002). Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Econ*, *11*(5), 415-30.

IsHak, W. W., Elbau, I., Ismail, A., Delaloye, S., Ha, K., Bolotaulo, N. I., … Wang, C. (2013). Quality of life in borderline personality disorder. *Harv Rev Psychiatry*, *21*(3), 138-50.

Jacob, G., & Arntz, A. (2003). Schema therapy for personality disordersâĂŤa review. *International Journal of Cognitive Therapy*, *6*, 171-185.

Jansson-Frojmark, M. (2014). The work and social adjustment scale as a measure of dysfunction in chronic insomnia: Reliability and validity. *Behav Cogn Psychother*, 1-13.

Kellogg, S. H., & Young, J. E. (2006). Schema therapy for borderline personality disorder. *J Clin Psychol*, *62*(4), 445-58.

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., … Walters, E. E. (2005). The world health organization adult ADHD Self-Report Scale (ASRS): a short screening scale for use in the general population. *Psychol Med*, *35*(2), 245-56.

Knapp, M. (1999). Economic evaluation and mental health: sparse past, fertile future? *J Ment Health Policy Econ*, *2*(4), 163-167.

Leichsenring, F., Leibing, E., Kruse, J., New, A. S., & Leweke, F. (2011). Borderline personality disorder. *The Lancet*, *377*(9759), 74-84.

Lobbestael, J., Arntz, A., Harkema-Schouten, P., & Bernstein, D. (2009). Development and psychometric evaluation of a new assessment method for childhood maltreatment experiences: the interview for traumatic events in childhood (ITEC). *Child Abuse Negl*, *33*(8), 505-17.

Lobbestael, J., van Vreeswijk, M., Spinhoven, P., Schouten, E., & Arntz, A. (2010). Reliability and validity of the short schema mode inventory (SMI). *Behav Cogn Psychother*, *38*(4), 437-58.

Lobbestael, J., Van Vreeswijk, M. F., & Arntz, A. (2008). An empirical test of schema mode conceptualizations in personality disorders. *Behav Res Ther*, *46*(7), 854-60.

Lockwood, G., & Perris, P. (2012). A new look at core emotional needs. In M. van Vreeswijk, J. Broersen, & M. Nadort (Eds.), *The wiley-blackwell handbook of schema therapy: Theory, research, and practice* (p. 41-66). John Wiley & Sons, Ltd.

Lynch, T. R., Trost, W. T., Salsman, N., & Linehan, M. M. (2007). Dialectical behavior therapy for borderline personality disorder. *Annu Rev Clin Psychol*, *3*, 181-205.

Manca, A., Hawkins, N., & Sculpher, M. J. (2005). Estimating mean qalys in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Econ*, *14*(5), 487-96.

Manca, A., Sculpher, M. J., & Goeree, R. (2010). The analysis of multinational cost-effectiveness data for reimbursement decisions: a critical appraisal of recent methodological developments. *Pharmacoeconomics*, *28*(12), 1079-96.

Manca, A., & Willan, A. R. (2006). 'lost in translation': accounting for between-country differences in the analysis of multinational cost-effectiveness data. *Pharmacoeconomics*, *24*(11), 1101-19.

Mataix-Cols, D., Cowley, A. J., Hankins, M., Schneider, A., Bachofen, M., Kenwright, M., … Marks, I. M. (2005). Reliability and validity of the work and social adjustment scale in phobic disorders. *Compr Psychiatry*, *46*(3), 223-8.

McCrone, P. (2011). Mental health economics: current methodological issues. *Epidemiol Psychiatr Sci*, *20*(3), 239-43.

McMain, S. F., Links, P. S., Gnam, W. H., Guimond, T., Cardish, R. J., Korman, L., & Streiner, D. L. (2009). A randomized trial of dialectical behavior therapy versus general psychiatric management for borderline personality disorder. *American Journal of Psychiatry*.

Mohr, D. C., Spring, B., Freedland, K. E., Beckner, V., Arean, P., Hollon, S. D., … Kaplan, R. (2009). The selection and design of control conditions for randomized controlled trials of psychological interventions. *Psychother Psychosom*, *78*(5), 275-84.

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. H. (2002). The work and social adjustment scale: a simple measure of impairment in functioning. *Br J Psychiatry*, *180*, 461-4.

Nadort, M., Arntz, A., Smit, J. H., Giesen-Bloo, J., Eikelenboom, M., Spinhoven, P., … van Dyck, R. (2009). Implementation of outpatient schema therapy for borderline personality disorder with versus without crisis support by the therapist outside office hours: A randomized trial. *Behaviour Research and Therapy*, *47*(11), 961-973.

O'Hagan, A., & Luce, B. R. (2003). *A primer on Bayesian statistics in health economics and outcomes research.*

Pang, F. (2002). Design, analysis and presentation of multinational economic studies: the need for guidance. *Pharmacoeconomics*, *20*(2), 75-90.

Perris, P., Young, J., Lockwood, G., Arntz, A., & Farrell, J. (2008). *Emotional core needs inventory.* www.cbti.se.

Perry, J. C., Banon, E., & Ianni, F. (1999). Effectiveness of psychotherapy for personality disorders. *Am J Psychiatry*, *156*(9), 1312-21.

Perseius, K. I., Andersson, E., Asberg, M., & Samuelsson, M. (2006). Health-related quality of life in women patients with borderline personality disorder. *Scand J Caring Sci*, *20*(3), 302-7.

Rijkeboer, M. M., van den Bergh, H., & van den Bout, J. (2005). Stability and discriminative power of the young schema-questionnaire in a Dutch clinical versus non-clinical population. *J Behav Ther Exp Psychiatry*, *36*(2), 129-44.

Samuels, J. (2011). Personality disorders: epidemiology and public health issues. *Int Rev Psychiatry*, *23*(3), 223-33.

Scharfe, E., & Bartholomew, K. (1994). Reliability and stability of adult attachment patterns. *Personal Relationships*, *1*, 23 - 43.

Schulman, K., Burke, J., Drummond, M., Davies, L., Carlsson, P., Gruger, J., … Glick, H. (1998). Resource costing for multinational neurologic clinical trials: methods and results. *Health Econ*, *7*(7), 629-38.

Sculpher, M. J., Pang, F. S., Manca, A., Drummond, M. F., Golder, S., Urdahl, H., … Eastwood, A. (2004). Generalisability in economic evaluation studies in healthcare: a review and case studies. *Health Technol Assess*, *8*(49), iii-iv, 1-192.

Soeteman, D. I., Hakkaart-van Roijen, L., Verheul, R., & van Busschbach, J. J. (2008). The economic burden of personality disorders in mental health care. *Journal of Clinical Psychiatry*, *69*(2), 259-265.

Soeteman, D. I., Verheul, R., & van Busschbach, J. J. (2008). The burden of disease in personality disorders: diagnosis-specific quality of life. *J Pers Disord*, *22*(3), 259-68.

van Asselt, A. D., Dirksen, C. D., Arntz, A., Giesen-Bloo, J. H., van Dyck, R., Spinhoven, P., … Severens, J. L. (2008). Out-patient psychotherapy for borderline personality disorder: cost-effectiveness of schema-focused therapy v. transference-focused psychotherapy. *Br J Psychiatry*, *192*(6), 450-7.

van Asselt, A. D., Dirksen, C. D., Arntz, A., & Severens, J. L. (2007). The cost of borderline personality disorder: societal cost of illness in BPD-patients. *Eur Psychiatry*, *22*(6), 354-61.

Veenhoven, R. (2008). *http://worlddatabaseofhappiness.eur.nl/hap_quer/hqs_fp.htm.*

www.euroqol.org. (2014).

Young, J. E. (1998). *Young schema questionnaire short form*. New York: Cognitive Therapy Centre.

Young, J. E., Arntz, A., Atkinson, T., Lobbestael, J., Weishaar, M., van Vreeswijk, M., & Klokman, J. (2007). *The schema mode inventory*. New York: Schema Therapy Institute.

Chapter 5

Economic evaluation of group schema therapy for borderline personality disorder: a multicentre, randomized controlled trial in the Netherlands

Pim Wetzelaer[1], Arnoud Arntz[1,2], Marleen Rijkeboer[1] & Silvia Evers[3,4]

[1]Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands
[2]Department of Clinical Psychology, University of Amsterdam, Amsterdam, The Netherlands
[3]Department of Health Services Research, CAPHRI Care and Public Health Research Institute, Faculty of Health Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands
[4]Centre of Economic Evaluation, Trimbos Institute (Netherlands Institute of Mental Health and Addiction), Utrecht, The Netherlands

Chapter 6

Bayesian multilevel net benefit regression for longitudinal cost-effectiveness data

Pim Wetzelaer[1], Joran Jongerling[2,3], Silvia Evers[4,5] & Arnoud Arntz[1,6]

[1]Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands
[2]Department of Methodology and Statistics, Faculty of Social and Behavioural Sciences, Utrecht University, Utrecht, The Netherlands
[3]Department of Pedagogical and Educational Sciences, Faculty of Social Sciences, Erasmus University, Rotterdam, The Netherlands
[4]Department of Health Services Research, CAPHRI School of Public Health and Primary Care, Faculty of Health Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands
[5]Trimbos Institute (Netherlands Institute of Mental Health and Addiction), Utrecht, The Netherlands
[5]Department of Clinical Psychology, University of Amsterdam, Amsterdam, The Netherlands

6.1 Abstract

Multilevel modelling is an appropriate solution for economic evaluations based on longitudinal randomized controlled trials (RCTs) in which calculation of total costs per individual patient is jeopardized by missing data. When applied in a Bayesian context, the estimates of model parameters can be used to calculate the probability that an intervention is cost-effective relative to another. In this paper we demonstrate how to apply Bayesian multilevel modelling to longitudinal cost-effectiveness data from (multicentre) randomized clinical trials as an extension of the net benefit regression framework. Using an empirical dataset we illustrate how it can be tested whether the development of net benefit over time is best described by 1) autoregressive models, 2) latent growth curve models, 3) quadratic growth curve models, or 4) autoregressive latent trajectory models. Cost-effectiveness acceptability curves were directly constructed using the parameter estimates from the best fitting model and based on all available data, including cases with missing values. We conclude that Bayesian multilevel net benefit regression is a useful approach for the analysis of longitudinal cost-effectiveness data.

6.2 Introduction

Randomized clinical trials (RCTs) on interventions for complex health problems often require extensive follow-up time periods before their full impact is revealed. These follow-up time periods may exceed the timespan for which it is still reasonable to rely on patient recall. Therefore, when longitudinal data are collected through patient report (e.g., interviews or questionnaires) assessments are usually performed repeatedly over time. An advantage of this repeated assessment is that it can show the (individual differences in) participants' development over time. This can provide important insight into the dynamics of the health problems under study that would otherwise remain obscured. However, repeated assessment may also cause a specific methodological challenge. It inherently creates opportunities for missing data to occur: whenever a trial participant misses a planned assessment, this will create a missing value in the dataset.

In general, the loss of information due to missing data adds uncertainty to the results that should not be ignored. Otherwise, results might be (seriously) biased. A second drawback of missing data is more specific for economic evaluations: when a patient is repeatedly assessed, missing data for one or more assessments jeopardizes the calculation of the total costs that were incurred for this patient over the course of the complete follow-up time period. Even when for a patient data is available for all but a single assessment, there will be a missing value for the total costs for that patient. This presents a researcher

who wants to perform an analysis based on total costs per patient with two options. The first is to completely remove a patient from the analysis, even if he/she missed only a single assessment. Total costs for that patient will then be treated as a missing value. The second option is to impute the missing values for individual assessments so that the total costs for that patient can be estimated. This would imply that additional procedures need to be performed for imputation, which can be rather complex. A more direct and efficient use of the data would be using all *available* patient data. So, if a patient missed one (or more) assessments, then all the data that *is* available from this patient (i.e., the scores on the assessments that the patient did not miss) could still be used for parameter estimation. This can be done using multilevel modelling.

Multilevel modelling was developed to analyse hierarchical data in which observations are 'nested' within overarching units (Hox, 1995). Examples of such nested data are repeated assessments, in which the consecutive observations are nested within individuals, and multicentre trials, in which participants are nested in treatment centres. Multilevel analysis efficiently uses all available data when estimating parameters. For example, assume that we assess a sample of 10 individuals on 5 different time points, and that one of these individuals has missing values at the third and fifth assessment. The data for the other 9 individuals are complete. With these data, multilevel analysis will estimate the mean score at time points 1, 2, and 4, using all 10 individuals in the sample. To estimate the mean scores at time points 3 and 5, multilevel analysis will use the *observed* scores of the 9 individuals with complete data, and the *expected* scores at these time points for the individual with the missing data, where these expected scores are determined based on his/her observed scores at time points 1, 2, and 4. This way, all information provided by the individuals in the sample is used, while the fact that there is missing data will show up in larger standard errors for the estimates at time point 3 and 5 (because the estimates for those time points are based on fewer *observed* scores than the estimates at time point 1, 2, and 4). This way of handling missing data implies that multilevel analysis will easily produce unbiased results (under the assumption that the missingness is either random or conditional on other observed variables in the model, see Little & Rubin, 2014) when data includes cases with missing values. Multilevel modelling thus has the ability to account for hierarchical data structures, and can easily deal with the missing data that are likely to occur with longitudinal research designs in which data collection is based on patient report. Therefore, it is ideally suited for analyses of data from economic evaluations performed alongside longitudinal (multicentre) RCTs.

Economic evaluations provide evidence on the value-for-money of health care interventions (Drummond et al., 2015). Specifically, patient-level data on health outcomes (e.g., percentage of recovery) are combined with economic data on resource use in what is called a cost-effectiveness analysis (CEA). A convenient framework for CEA is net

benefit regression (NBR; Hoch et al., 2002). In this framework, the net benefit (NB) is calculated for each trial participant by subtracting the costs incurred by this individual from the amount that society is willing to pay for his or her health outcome, and subsequently used as the dependent variable in a regression analysis. Differences in mean net benefit between different treatments can be determined by also adding the type of treatment as a predictor in the regression model. A positive difference in mean net benefit of one treatment over another indicates the relative cost-effectiveness of that treatment.

Ultimately, the goal of a CEA is to give probabilistic statements about which intervention is the most cost-effective. However, with the standard (frequentist) analysis methods available in popular statistical software packages such as SPSS, STATA, and SAS, this is not possible. The reason for this is that these methods can only be used for null hypothesis testing, the method of inference used in classical statistics. In the present context this would imply calculating the probability of finding the observed sample data when assuming the intervention of interest is not cost-effective. This probability is not the same as the probability that an intervention is cost-effective, and importantly, the one cannot be determined from the other. A solution to this problem is provided by Bayesian statistics. Bayesian statistics is a subset of the field of statistics that has a different approach to probability and statistical inference than classical statistics. Due to these differences, the outcome of a Bayesian analysis does not consist of parameter-estimates and their corresponding standard errors, but of entire distributions for the parameters under study. These distributions, called posterior distributions, give the likelihood for different parameter values given the data, and can be used for statistical inference. The mean of the posterior distribution of a parameter can be used as a point estimate for example, while the standard deviation can be used as a measure equivalent to standard errors. Moreover, the likelihood of certain ranges of parameter values (e.g., differences in mean net benefit larger than 0) can be estimated by determining the proportions of the posterior distributions that fall within this range. Using these estimates, cost-effectiveness acceptability curves that show the probability that an intervention is cost-effective can be constructed for a range of different willingness-to-pay values.

In this article, we aim to demonstrate how Bayesian multilevel net benefit regression models can be applied to longitudinal cost-effectiveness data from (multicentre) RCTs. First, we describe how net benefit can be calculated for each individual and for each assessment in longitudinal studies for the case of a dichotomous (i.e., binary) variable for effectiveness that is only assessed once in combination with a variable for costs that is repeatedly assessed as well as the case of a continuous variable for effectiveness in combination with a variable for costs that are both repeatedly assessed. Next, we describe in more detail how to use the estimates of NBR model parameters for calculating the probability of relative cost-effectiveness. Subsequently, we introduce the set of multi-

level NBR models that were used for this paper and describe how they can be estimated. These sections are deliberately kept short and avoid the use of technical terms to facilitate readability of the paper. However, we do provide more detailed (statistical) descriptions of the models as well as their estimation using Bayesian methods in the appendix to this paper. Next, we explain how to determine which model from among the set of models provided fits the data best and how the output of this model is presented. Finally, we demonstrate the above using an empirical dataset from a longitudinal, multicentre RCT on psychotherapy for personality disorders. This dataset is characterized by a dichotomous (i.e., binary) variable for effectiveness (i.e., indicating whether or not a patient has recovered at the three-year follow-up assessment), societal costs that were repeatedly assessed during the same three year follow-up time period, and missing values.

A detailed manual is provided to accompany this article with explanations and step-by-step instructions for data preparation and running the models that are presented in this paper (including R commands and JAGS model code). Importantly, the primary focus of the paper and supporting documents are to explain and demonstrate the application of the models. The instructions provided are restricted to the steps that are at the least necessary in producing the results here presented. As such, they are not intended to be wholly sufficient for the guidance on the complete and proper analysis of a given dataset (e.g., no guidance is provided on how to perform additional sensitivity analyses to test assumptions regarding the randomness of missing data or prior distributions). However, some practical notes on the application of (Bayesian) regression modelling in general are provided where applicable.

6.3 Methods

6.3.1 Net benefit regression

In order to perform net benefit regression on longitudinal data, net benefits need to be calculated for each individual participant and for each of the repeated assessments. Subsequently, these net benefits can be used as the dependent variable in a regression framework. This contrasts with the conventional application of net benefit regression in which total net benefit for each individual participant is used as the dependent variable. We will demonstrate below how the net benefits can be calculated for longitudinal data in two cases. First, we describe how to do this when using a dichotomous outcome measure for effectiveness (i.e., recovery from borderline personality disorder (BPD) at follow-up, as used in the specific dataset by which we demonstrate how to apply our approach further on in this article). Second, we also describe how to calculate net benefits when a continuous outcome measure is used for effectiveness.

The calculation of net benefits for each assessment is based on the costs as measured at each of those assessments. These costs refer to the costs that were incurred between subsequent assessments, not the total costs that were incurred up until a specific assessment. Also the willingness-to-pay value (i.e., the amount of money used for the valuation of a participant's recovery in our example) is divided over the total number of assessments, so that it contributes equally to the net benefit for each assessment. It is furthermore assumed that all time periods between subsequent assessments have an equal duration.

Importantly, only the costs incurred once treatment has started need to be considered in an economic evaluation. This implies that the net benefit for the baseline assessment, although it is used for modelling the development of net benefit over time (i.e., to estimate the intercept), is not taken into account in the calculation of total mean net benefit per treatment. Therefore, the total willingness-to-pay value is divided by the total number of assessments minus one when calculating the net benefits for each individual participant and for each of the repeated assessments.

For longitudinal studies in which a dichotomous outcome measure is used for effectiveness, the net benefit (NB) can be determined for $i = 1, 2, \ldots, \mathrm{N}$ individuals (where N denotes the total sample size), and for $t = 1, 2, \ldots, \mathrm{T}$ assessments (where T is the total number of repeated assessments) using,

$$NB_{i,t} = \left(\frac{\lambda * E_i}{T-1}\right) - C_{i,t},$$

where E_i is a dichotomous variable indicating effectiveness of the treatment for patient i (in our empirical example $E_i = 0$ indicates no recovery at follow-up, and $E_i = 1$ indicates recovery at follow-up), λ is the willingness-to-pay value in euros, and $C_{i,t}$ are the costs for patient i incurred in the time period between assessment t and $t-1$. In longitudinal studies in which a continuous outcome measure is used as the effect variable, the net benefit can be determined using,

$$NB_{i,t} = \left(\frac{\lambda * (E_{i,t} - E_{i,t-1})}{T-1}\right) - C_{i,t},$$

where $E_{i,t}$ is a continuous variable indicating the effectiveness of the treatment for patient i at assessment t.

These net benefits are subsequently used as the dependent variable in the NBR models. In these models, conclusions about the relative cost-effectiveness of treatments are based on differences in the mean total net benefit between treatments: a positive difference in mean total net benefit for one treatment over another indicates the relative

cost-effectiveness of that treatment. The mean total net benefit of each treatment is calculated by first determining each treatment's mean net benefit at each assessment (using Bayesian estimation, see Appendix A) and subsequently summing the separate estimates for assessments $t = 2, \ldots, T$ (while excluding the baseline assessment as explained above).

6.3.2 Model specification

In this paper we use Bayesian methods to test for a specific dataset whether the development of net benefit over time is best described by 1) a first-order autoregressive (AR1) model (Jöreskog, 1970, 1979), 2) a latent growth curve (LGC) model (Bollen & Curran, 2004, 2006; Meredith & Tisak, 1990), 3) a quadratic growth curve (QGC) model, or 4) an autoregressive latent trajectory (ALT) model (Bollen & Curran, 2004; Curran & Bollen, 2001). A detailed statistical description of all the models that were used in this study is provided as an appendix to this paper (Appendix A). In short, the underlying rationale for each of the models is the following: AR1 models test if net benefit shows reversible change over time in the absence of systematic change, LGC models test for linear change in net benefit, QGC models test for curvilinear change, and ALT models combine a LGC model with an AR1 model, and can therefore test for linear change while accounting for the possibility of autocorrelation between consecutive observations obtained from the same individual. These longitudinal multilevel models are well suited for modelling the development of net benefit over time. Not only do they allow for the estimation of individual change trajectories (or growth curves) for net benefit over time, but for the estimation of inter-individual differences in these trajectories as well. This also implies that these models can be easily extended to account for differences at higher levels, such as differences between treatment centres in a multicentre RCT (see Appendix C for an example of a three level model).

All models were tested using both fixed and random parameters, and in all possible combinations thereof. For the present article, this was done for the purpose of demonstrating the full set of models. In practice however, it would be more sensible to select a subset of models consistent with the results of a preceding preliminary analysis of the data (i.e., using simple descriptive statistics and plots; see Appendix D). Following multilevel modelling terminology, random parameters are defined as consisting of both a fixed part (e.g., a mean intercept or slope) and a random part (e.g., individual deviations from the mean intercept or slope). This implies that when a parameter is included as a random effect, the fixed part of the parameter is always included in the model as well. In other words, an estimate of the fixed part of the random parameter is equivalent to the estimate of a fixed parameter for the same variable. This means that when estimating a random parameter for a given variable, the fixed parameter for the same variable is automatically

included as well. For the AR1 model and the ALT model, we did not estimate models with random AR-parameters, because the number of repeated assessments is too small for meaningful estimation of inter-individual differences in this parameter. For the AR1 model, the accompanying manual does provide code for data analysis with random AR-parameter. For the ALT model, no such code is provided because this model is usually not applied to data with enough repeated assessments for random AR-parameters.

For Bayesian analysis, one needs to specify prior distributions (representing prior beliefs about parameter values) and data distributions, both of which are given in Appendix A. In short, vague (or diffuse) priors were used for all models and all models were based on normally distributed data. Throughout the current article we limited our analyses to normally distributed net benefit data with just two-levels (observations nested in individuals). However, both data with non-normal distributions, and data with more than two levels can occur, in particular for economic evaluations performed in parallel to multicentre RCTs. Costs in economic evaluations often follow non-normal, skewed distributions due to the fact that costs are non-negative and whereas most patients typically incur relatively low costs, few(er) patients will incur high(er) costs. Since net benefit is in part a function of costs, when costs are skewed this will also affect the distribution of net benefit. Multicentre longitudinal RCTs, in which patients are assessed in several participating centres, lead to three-level data in which observations (level 1) are nested in patients (level 2), and patients are nested in treatment centres (level 3). Information on how to extend our approach to non-normally distributed data and/or nested data with three or more levels is provided in Appendix C.

6.3.3 Model estimation

The models described in this paper are estimated using JAGS, a program designed for analyzing Bayesian models using Markov Chain Monte Carlo (MCMC) estimation. Conveniently, JAGS interfaces with R, a programming environment well suited for the implementation of statistical techniques. We provide a detailed manual to accompany this article that explains the different steps necessary for the preparation of the data in R, the code that shows how the different models are implemented in JAGS as well as the R commands required to run the models in JAGS according to specific settings. These settings pertain to total number of iterations, the number of Markov chains that are run for each model, the 'burn-in' or the number of iterations at the beginning of the chain that are discarded and the thinning rate, which is a positive integer n that indicates that for every iteration of the model only the values of the nth iteration are used. This serves to reduce autocorrelation between values from successive iterations, which is needed in certain cases for successful parameter estimation. Initially, the models are estimated us-

ing 100,000 iterations, of which the first 5,000 are discarded (i.e., burn-in). A standard thinning rate of 1 is used. The Gelman-Rubin statistic (Gelman & Rubin, 1992) is used as a diagnostic criterion for convergence, with a cutoff value of 1.1. In addition, visual inspection of the 'trace plots' (plots that show the sampled values of parameters for each iteration) is used to confirm that convergence has been achieved (i.e., these trace plots appear as 'fat hairy caterpillars' in which sampled values scatter randomly about a stable mean value, which would be indicative for convergence, or they appear as 'thin curly snakes' when convergence has not yet been reached (Lunn et al., 2013)). In case convergence has not been reached after 100,000 iterations, additional iterations are run (using the 'update' function in JAGS) until convergence is reached. Also, convergence issues can often be resolved by better choice of starting values, or re-parameterising the model (e.g., ensure all covariates are centred). In case convergence cannot be reached at all, this suggests conflict between the data and model (which includes prior specification). Non-convergence means that the estimates from repeated iterations do not stabilize around a single mean value. As a consequence, the posterior distributions cannot be interpreted as a reliable reflection of the parameter values. Therefore, a different model should be used in case of non-convergence.

6.3.4 Model comparison

The different models are compared based on their deviance information criterion (DIC) values to determine which model fits the data best (Spiegelhalter et al., 2002). A description of this information criterion is presented in Appendix B. For now, it suffices to know that, similar to other information criteria like the AIC (Akaike, 1998) and BIC (Schwarz, 1978), the DIC can be seen as consisting of two parts; one part that measures model (mis)fit, and a second part that quantifies the dimensionality, or complexity, of a model. This implies that model selection based on the DIC is based on a trade-off between model fit and model complexity. If two models fit the data equally well, than the model with the lowest complexity (i.e., the lowest number of model parameters) will be selected. Lower values on the DIC imply a 'better' model fit and, as a rule of thumb, differences in DIC values larger than 5 are usually considered relevant. Once it is decided which model fits the data best, treatment condition is added as a predictor for the inter-individual variances in the intercept and the slope in order to estimate separate growth curves for each treatment. It is important to note that for a complete data analysis the DIC should never be used in isolation for model comparison. This is because in principle the possibility exists that all models under consideration have a poor fit to the data. Therefore, additional model checking (e.g., residual checks, predictive checks (Lunn et al., 2013; Gelman et al., 2014)) would be required when doing a complete data analysis.

In addition, it is important to note that model checking and comparison are complicated by hierarchical structure (Lunn et al., 2013).

6.3.5 Presentation of the results

The change in mean NB is presented using separate growth curves for each treatment, that is, curves that show each treatment's model predicted mean NB for the individual assessments. These curves are constructed using our best fitting model and a range of different values for willingness-to-pay. The information on relative cost-effectiveness between treatments can be plotted in cost-effectiveness acceptability curves (CEACs). CEACs plot the probability of relative cost-effectiveness for one treatment over another, based on incremental mean total net benefit, for a range of different willingness-to-pay values λ. Specifically, willingness-to-pay per recovered patient is given on the x-axis of these CEACs, whereas the probability of relative cost-effectiveness is given on the y-axis. Growth curves as well as CEACs were estimated using the following range of willingness-to-pay values (λ; in euros): 0; 2,500; 5,000; 7,500; 10,000; 12,500; 15,000; 17,500; 20,000; 30,000; 37,500; 40,000; 60,000 and 80,000.

6.3.6 Empirical dataset

To demonstrate how to apply the Bayesian multilevel models and choose between the different models in practice, the suggested approach is illustrated using a dataset from a multicentre RCT on psychotherapy for personality disorders (PDs). The study protocol and results have been described in detail elsewhere (Bamelis et al., 2012, 2014, 2015). In short, this RCT included 320 patients with various PDs recruited from twelve mental health centres in the Netherlands, who were randomized over three treatment conditions: schema therapy (ST; n=145), clarification-oriented psychotherapy (COP; n=41) and treatment as usual (TAU; n=134). Overall, the dataset contained 22% missing data (from 0% at baseline to 32% at follow-up). Importantly, it is assumed that data are missing at random (MAR) or missing completely at random (MCAR; Little & Rubin, 2014). A demonstration of the additional sensitivity analyses that would be required when there are reasons to question the plausibility of this assumption (Molenberghs et al., 2014) are beyond the scope of this paper. Societal costs were assessed at six points in time in total. After an initial baseline assessment (and subsequent start of treatment), societal costs were assessed every six months for two years. A final follow-up assessment took place one year later (i.e., at three years), at which point it was also assessed whether or not the patient had recovered from BPD. We have coded an additional time point t_6, containing missing values for the net benefit of each patient, to preserve six month intervals between

time points. Therefore, growth curves show seven time points in total. Data on patients' recovery from PD at three-year follow-up and their total societal costs incurred in the six months before each assessment were used to calculate NB for each individual and for each of the six assessments.

6.4 Results

The DIC values of all models that were tested are listed in Table 6.1. In addition, the *pD*-values for all the models, which are indicative of model complexity (higher values indicate more complexity), are given in this table. In case a model did not converge, this is also listed. These models are either too complex for the amount of information present in the data, or they have a really poor fit to the data. A QGC model with random intercept, random slope, and fixed quadratic term had the lowest DIC value and was therefore the best fitting model (see Table 6.1). Subsequently, type of treatment was added as a predictor for the inter-individual variances in the intercept and the slope to enable estimation of separate growth curves for each treatment. These growth curves are shown in Fig 6.1. This figure shows that, although the NB of the COP treatment seems to be structurally lower than that of the other two treatments, the development of mean NB is quite similar for all three treatments. The parameter estimates and their corresponding credibility intervals (listed in Table 6.2) show a similar picture, with the credibility intervals of the different treatments showing considerable overlap. However, since the credibility intervals are quite wide, the absence of substantial differences between conditions may be due to low statistical power. Next, the probability of relative cost-effectiveness (i.e., the probability that one treatment has a higher mean NB than the others) was determined. The corresponding CEACs in Fig 6.2 show that the ST treatment is the most cost-effective of the three and the COP treatment the least cost-effective. Based on these results, the ST treatment would be recommended over the other two. However, it is important to keep in mind that our regression analysis did not include any covariates (other than a treatment indicator variable) that might lead to different results. We also performed an analysis on the basis of the same model, but now using lognormal and gamma distributed data. The CEACs resulting from this analysis are presented in Fig 6.3 and Fig 6.4. Again, the ST treatment is the most cost-effective of the three treatments in both analyses. However, the CEACs for the COP and TAU treatments have slightly changed in comparison to the analysis using normally distributed data.

Table 6.1: The DIC values and *pD* values for the different Bayesian multilevel models.

Model family	Variant	DIC[1]	*pD*[2]
AR	Fixed Mean	29767.90	61.30
	Random Mean	Did not converge	
LGC	Fixed Intercept and Slope	30152.20	3.00
	Random Intercept, Fixed Slope	29944.30	497.20
	Fixed Intercept, Random Slope	30335.00	302.80
	Random Intercept and Slope	29737.70	984.40
QGC	Fixed Intercept, Slope and Quadratic Term	30154.20	4.00
	Random Intercept, Fixed Slope and Quadratic Term	29959.40	511.20
	Random Slope, Fixed Intercept and Quadratic Term	30348.10	313.30
	Random Quadratic Term, Fixed Intercept and Slope	Did not converge	
	Random Intercept and Slope, Fixed Quadratic Term	29728.00	972.90
	Random Intercept and Quadratic Term, Fixed Slope	30196.30	990.30
	Random Slope and Quadratic Term, Fixed Intercept	30589.20	588.40
	Random Intercept, Slope, and Quadratic Term	Did not converge	
ALT (constrained)	Fixed Alpha, Beta, and AR	29738.80	56.50
	Random Alpha, Fixed Beta and AR	Did not converge	

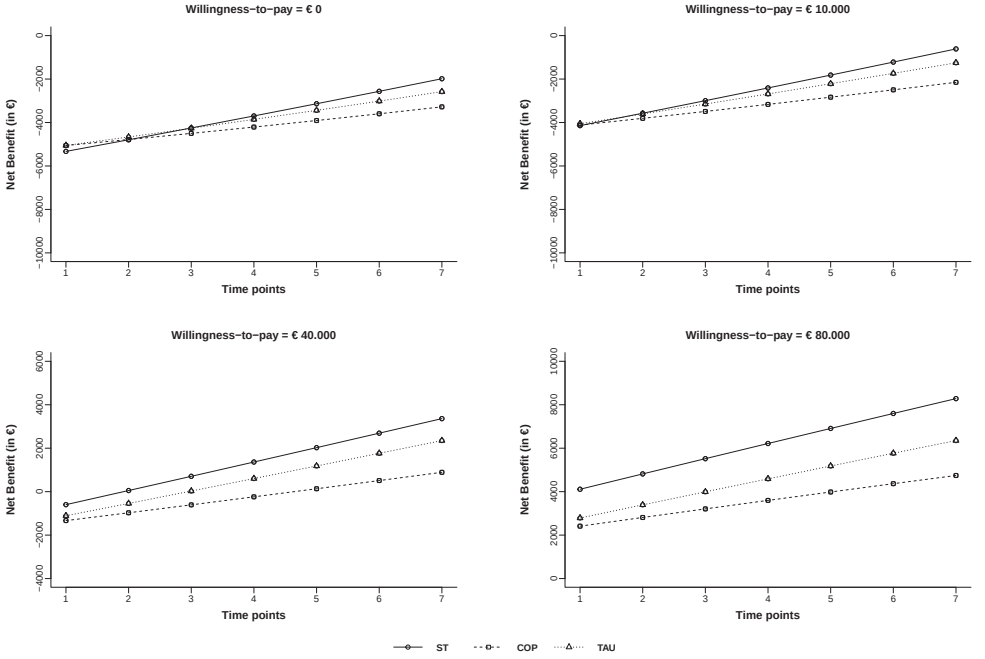
Continued on next page

Table 6.1 – continued from previous page

Model family	Variant	DIC[1]	pD^2
	Random Beta, Fixed Alpha and AR	30369.50	973.30
	Random Alpha and Beta, Fixed AR	Did not converge	
ALT (LGC with AR(1) Errors)	Fixed Intercept, Slope, and AR	29739.30	56.80
	Random Intercept, Fixed Slope and AR	Did not converge	
	Random Slope, Fixed Intercept and AR	Did not converge	
	Random Intercept and Slope, Fixed AR	Did not converge	

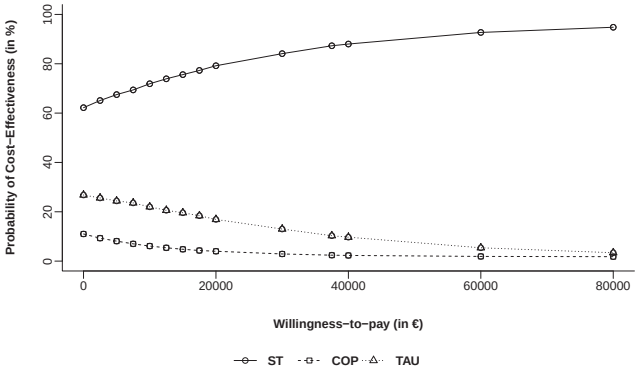
Notes:
[1] Lower values for the DIC indicate better model fit.
[2] Lower values for *pD* indicate lower model complexity.

Figure 6.1: Growth curves for the different treatments.



Notes: Based on the QGC model with random intercept and slope, fixed quadratic term and predictors for treatment and using willingness-to-pay values of € 0 (upper left panel), € 10.000 (upper right panel), € 40.000 (lower left panel) and € 80.000 (lower right panel). These growth curves were constructed using posterior means. For an indication of the uncertainty to these predictions, please see the 95% credibility intervals in Table 6.2.

Figure 6.2: Cost-effectiveness acceptability curves.



Notes: These CEACs are based on the QGC model with random intercept, random slope, fixed quadratic term, and predictors for treatment condition, and using normally distributed net benefit data.

Table 6.2: Parameter estimates for the different willingness-to-pay values.

WTP	Group	Parameter estimates			95% Credibility interval		
		Intercept	Slope	Quadratic term	Intercept	Slope	Quadratic term
0	TAU	-5054.85	385.40	4.55	-6194.69 - -3932.09	-19.10 - 787.93	-42.03 - 53.56
	ST	-5329.62	530.08	4.55	-7976.38 - -2689.58	-281.04 - 1328.92	-42.03 - 53.56
	COP	-5056.56	268.84	4.55	-8381.39 - -1669.64	-724.48 - 1245.76	-42.03 - 53.56
2500	TAU	-4794.92	397.20	5.14	-5883.37 - -3694.39	7.83 - 800.26	-41.66 - 52.93
	ST	-5019.95	533.23	5.14	-7633.61 - -2420.82	-251.05 - 1344.75	-41.66 - 52.93
	COP	-4843.91	275.55	5.14	-8178.35 - 1492.57	-707.35 - 1247.03	-41.66 - 52.93
5000	TAU	-4573.07	423.11	4.08	-5678.40 - -3416.32	16.84 - 824.96	-43.55 - 52.61
	ST	-4742.44	551.01	4.08	-7404.85 - -2050.64	-263.35 - 1350.04	-43.55 - 52.61
	COP	-4606.87	286.22	4.08	-7984.63 - -1206.82	-712.63 - 1284.34	-43.55 - 52.61
7500	TAU	-4315.34	432.84	4.06	-5417.44 - -3183.88	45.15 - 826.23	-42.65 - 53.36
	ST	-4439.69	558.64	4.06	-7028.16 - -1780.44	-22.48 - 1361.72	-42.65 - 53.36
	COP	-4376.14	292.08	4.06	-7738.85 - -972.21	-681.90 - 1247.90	-42.65 - 53.36
10000	TAU	-4056.92	446.74	3.44	-5153.76 - -2938.37	51.92 - 841.89	-45.48 - 51.21
	ST	-4142.90	567.56	3.44	-6783.01 - -1532.48	-223.30 - 1357.37	-45.48 - 51.21
	COP	-4118.67	307.27	3.44	-7456.44 - -795.72	-659.61 - 1281.23	-45.48 - 51.21
12500	TAU	-3820.85	463.42	3.73	-5000.81 - -2676.27	55.74 - 864.52	-42.77 - 51.46
	ST	-3854.83	576.56	3.73	-6590.78 - -1194.34	-216.89 - 1383.69	-42.77 - 51.46
	COP	-3915.13	314.23	3.73	-7312.58 - -476.91	-678.56 - 1301.56	-42.77 - 51.46
15000	TAU	-3570.82	477.74	3.51	-4704.79 - -2473.86	80.40 - 868.02	-43.64 - 49.61
	ST	-3570.82	587.02	3.51	-6256.30 - -908.65	-218.56 - 1369.07	-43.64 - 49.61
	COP	-3702.26	324.45	3.51	-7178.54 - -369.96	-653.31 - 1286.63	-43.64 - 49.61
17500	TAU	-3331.68	493.11	2.66	-4461.26 - -2190.55	92.12 - 895.50	-45.76 - 52.39
	ST	-3262.72	598.81	2.66	-5977.80 - -574.23	-190.78 - 1400.65	-45.76 - 52.39

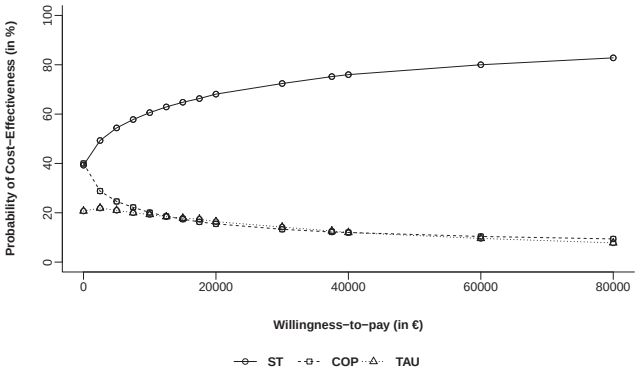
Continued on next page

Table 6.2 – continued from previous page

WTP	Group	Parameter estimates			95% Credibility interval		
		Intercept	Slope	Quadratic term	Intercept	Slope	Quadratic term
	COP	-3432.81	325.53	2.66	-6825.27 - 27.44	-695.13 - 1301.52	-45.76 - 52.39
20000	TAU	-3096.99	502.83	3.03	-4247.59 - -1976.87	107.55 - 897.78	-43.82 - 52.08
	ST	-2972.77	602.26	3.03	-5670.04 - -241.18	-202.79 - 1399.11	-43.82 - 52.08
	COP	-3185.70	326.85	3.03	-6676.35 - 376.67	-636.60 - 1288.56	-43.82 - 52.08
30000	TAU	-2083.47	539.39	2.17	-3274.63 - -883.20	148.66 - 942.53	-45.32 - 48.69
	ST	-1800.27	637.18	2.17	-4685.45 - 1047.31	-170.55 - 1445.21	-45.32 - 48.69
	COP	-2246.44	348.72	2.17	-5811.32 - 1443.04	-621.65 - 1324.90	-45.32 - 48.69
37500	TAU	-1351.97	565.48	0.72	-2603.62 - -95.82	171.92 - 974.64	-49.76 - 48.51
	ST	-917.74	659.44	0.72	-3833.65 - 2011.63	-129.29 - 1469.21	-49.76 - 48.51
	COP	-1520.92	361.37	0.72	-5345.05 - 2322.34	-610.03 - 1371.38	-49.76 - 48.51
40000	TAU	-1109.61	564.94	1.98	-2358.90 - 141.60	163.78 - 961.66	-46.11 - 49.73
	ST	-598.67	648.10	1.98	-3568.56 - 2377.72	-148.70 - 1458.23	-46.11 - 49.73
	COP	-1332.97	358.70	1.98	-5152.27 - 2512.55	-615.88 - 1313.41	-46.11 - 49.73
60000	TAU	829.95	604.01	-.38	563.78 - 2172.70	201.62 - 997.09	-47.11 - 47.00
	ST	1751.37	691.05	-.38	1473.27 - 5009.59	-114.23 - 1479.67	-47.11 - 47.00
	COP	580.46	376.73	-.38	3737.85 - 4875.24	-602.59 - 1360.66	-47.11 - 47.00
80000	TAU	2787.09	607.17	-2.19	1207.36 - 4411.75	189.40 - 1007.01	-50.81 - 47.90
	ST	4110.16	708.44	-2.19	332.30 - 7920.81	-129.53 - 1506.83	-50.81 - 47.90
	COP	2411.55	401.98	-2.19	-2329.35 - 7262.98	-580.94 - 1379.99	-50.81 - 47.90

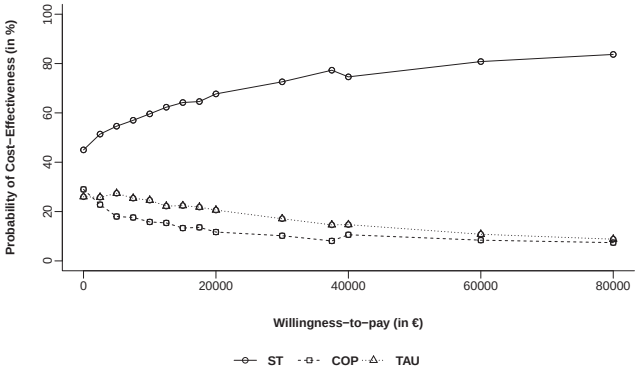
Notes: Parameter estimates and 95% credibility intervals for the QGC model with random intercept and slope, fixed quadratic term, and treatment as a predictor for inter-individual variability in the intercept and slope for a range of different willingness-to-pay values (WTP) values.

Figure 6.3: Cost-effectiveness acceptability curves using lognormal distributed data.



Based on the QGC model with random intercept, random slope, fixed quadratic term, and predictors for treatment condition, and using lognormal distributed data.

Figure 6.4: Cost-effectiveness acceptability curves using gamma distributed data.



Based on the QGC model with random intercept, random slope, fixed quadratic term, and predictors for treatment condition, and using gamma distributed data.

6.5 Discussion

In this paper, we have presented Bayesian multilevel modelling as an extension to the NBR framework because it has specific advantages when analyzing longitudinal cost-effectiveness data. First, it allows to specifically account for the hierarchical data structure inherent in repeated assessments and/or multicentre RCTs. Second, it efficiently makes use of all available data, including cases with missing values, and therefore renders additional procedures for imputation of missing data unnecessary. Third, the use of Bayesian methods allows estimation of the probability of cost-effectiveness for interventions. Fourth, it can flexibly account for non-normal data distributions. Fifth, it is transparent in terms of model specification. We believe the approach outlined in this paper provides an efficient way to perform a CEA on longitudinal data from (multicentre) RCTs from within a single coherent framework; the best fitting model is systematically determined out of a complementary set of models, then predictors for treatment condition are added and subsequently parameter estimates are directly used to produce growth curves and CEACs.

It is important to note that the aim of this study was to demonstrate a Bayesian multilevel approach to the analysis of longitudinal cost-effectiveness data, not to perform a re-analysis of the data from our empirical example. However, after including the same covariates as in the analysis reported in the original economic evaluation based on these data, both analyses lead to the same conclusion (just as the original base case analysis that did not use a regression approach, but was based on a bootstrap simulation) that ST is the superior treatment in terms of cost-effectiveness. Also, we considered that a further explanation and demonstration of any additional analyses that are required in practice for a complete data analysis (e.g., analyses for testing the sensitivity to the prior distributions that were used or sensitivity analyses when the assumption regarding randomness of missing data is implausible) were beyond our aim of demonstrating the application of the presented models to an empirical dataset.

Throughout the article we have limited our analyses to normally distributed net benefit data, and provide explanation of how to perform similar analyses based on non-normal (i.e., gamma and log-normal) distributions in the appendix to this article. Since net benefit is a function of both costs, which are often skewed to the right, and effects, the option to model also non-normally distributed net benefit data is useful in this context. In a follow-up study, we will investigate a two-dimensional approach to net benefit regression for longitudinal data, in which different distributions can be specified for costs and effects. Net benefits can then be calculated based on their joint posterior distribution. This will allow a closer approximation to the true underlying distribution of net benefits than the one-dimensional approach that we outlined in this article.

Complementary to this paper we provide a detailed manual that explains step-by-step how our approach can be implemented, including instructions for preparation of the data, model code and the R commands required to run the models. Furthermore, both in the appendix to this paper as well as in the manual we explain how the approach outlined in this paper can be extended to accommodate skewed data distributions (i.e., gamma and lognormal distributions) and nested data with three or more levels. We hope this will facilitate the use of our approach by applied researchers. For the future, we are furthermore planning to develop an R package that implements the methods outlined in this paper and that can be conveniently used as a tool for longitudinal cost-effectiveness analyses.

Bayesian multilevel modelling for the analysis of longitudinal cost-effectiveness data as an extension to the NBR framework holds much promise for applied researchers in the field of health economics specifically. Nonetheless, the abovementioned advantages of Bayesian multilevel modeling may also more generally appeal to applied researchers facing similar challenges in other research fields.

6.6 Acknowledgements

We are thankful for the comments and suggestions provided by two anonymous reviewers that were of great help to improve the paper. This study was supported by grants from the Netherlands Organization for Health Research and Development (ZonMW; 80-82310-97-12142) and Fonds Psychische Gezondheid (2008-6350), which were obtained by Arnoud Arntz.

6.7 References

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike* (p. 199-213). Springer.

Bamelis, L. L., Arntz, A., Wetzelaer, P., Verdoorn, R., & Evers, S. M. (2015). Economic evaluation of schema therapy and clarification-oriented psychotherapy for personality disorders: a multicenter, randomized controlled trial. *J Clin Psychiatry*, *76*(11), e1432-40.

Bamelis, L. L., Evers, S. M., & Arntz, A. (2012). Design of a multicentered randomized controlled trial on the clinical and cost effectiveness of schema therapy for personality disorders. *BMC public health*, *12*(1), 75.

Bamelis, L. L., Evers, S. M., Spinhoven, P., & Arntz, A. (2014). Results of a multicenter randomized controlled trial of the clinical effectiveness of schema therapy for personality disorders. *Am J Psychiatry*, *171*(3), 305-22.

Bollen, K. A., & Curran, P. J. (2004). Autoregressive latent trajectory (alt) models a synthesis of two traditions. *Sociological Methods & Research*, *32*(3), 336-383.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective* (Vol. 467). John Wiley & Sons.

Curran, P. J., & Bollen, K. A. (2001). The best of both worlds: Combining autoregressive and latent curve models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (p. 105-136). Washington, DC: American Psychological Association.

Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes*. Oxford university press.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2). Taylor & Francis.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 457-472.

Hoch, J. S., Briggs, A. H., & Willan, A. R. (2002). Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Econ*, *11*(5), 415-30.

Hox, J. J. (1995). *Applied multilevel analysis*. TT-publikaties Amsterdam.

Jöreskog, K. G. (1970). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology*, *23*, 121-145.

Jöreskog, K. G. (1979). Statistical estimation of structural models in longitudinal-developmental investigations. In J. Nesselroade & P. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (p. 303-352). New York: Academic Press.

Little, R. J., & Rubin, D. B. (2014). *Statistical analysis with missing data*. John Wiley & Sons.

Lunn, D., Jackson, C., Best, N., Thomas, A., & D., S. (2013). *The BUGS book: A practical introduction to Bayesian analysis*. Chapman and Hall.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*(1), 107-122.

Molenberghs, G., Fitzmaurice, G., Kenward, M. G., Tsiatis, A., & Verbeke, G. (2014). *Handbook of missing data methodology*. CRC Press.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461-464.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583-639.

Chapter 7

General discussion

7.1 Introduction

This thesis presents a collection of studies that were performed to advance the assessment and analysis of the cost-effectiveness of psychotherapy for borderline personality disorder. In Chapter 2 an overview of the current evidence base was given by means of a systematic review. Next, in Chapter 3 a model-based economic evaluation method for the synthesis of empirical evidence for cost-effectiveness and budget impact analyses was presented. Furthermore, in order to assess the (clinical and) cost-effectiveness of group schema therapy, a trial-based economic evaluation was performed and described in Chapters 4 and 5. Finally, Chapter 6 encompassed a methodological study into Bayesian multilevel models for the analysis of longitudinal cost-effectiveness data as an extension of the net benefit regression framework.

The present and concluding chapter discusses the different studies in general. First, the studies are integrated. Each study is shortly summarized in terms of its aims and findings and put into perspective relative to the other studies. The specific crossties and mutual complementarities between the studies are highlighted. This is to illustrate a 'bigger picture' as their overall sum of parts. Next, considerations are discussed that correspond to the methodologies underlying the studies. Based on the insights that follow from this, the prospects and recommendations for the further advancement of research in the future are addressed subsequently. Then the implications for policy and practice are discussed, followed last by a general conclusion of this thesis.

7.2 Integration

The studies that were performed for this thesis represent an amalgam of different research aims and methods. In addition to a short summary of the aims and findings of the studies, in this section each study is put into perspective relative to the other studies. Moreover, this section highlights the links between the studies and their findings. This is to facilitate an appreciation of the cohesion between the studies and how they combine into a larger whole. Furthermore, it provides a context against which the studies can be interpreted in terms of their added value.

The systematic review in Chapter 2 aimed to provide an overview of economic evaluation studies on psychotherapy for personality disorders. Most of the included studies indicated that psychotherapy for personality disorders is cost-effective. Yet, it was also noted that most studies did not include all (potentially) relevant societal costs, and that several studies did not use QALYs as an outcome measure.

For the study in Chapter 3, the clinical focus was narrowed relative to the systematic

review in the preceding chapter. Instead of focusing on the various psychotherapeutic interventions that have been studied for the treatment of different types of personality disorders in various settings, in Chapter 3 the focus is on specialized, outpatient psychotherapy for BPD. This was to allow an analysis of the cost-effectiveness and budget impact specifically for the types of treatment that are currently recommended for this disorder (as well as other cluster A or B personality disorder types) in the Netherlands and which are preferably delivered in an outpatient setting (i.e., unless otherwise indicated).

A method was presented in Chapter 3 for the synthesis of a wealth of information from studies with different designs and outcome measures, including empirical evidence based on information on health care costs, or resource use, and quality of life. Therefore, while the scope regarding diagnosis and treatment types was narrower, the scope regarding the inclusion of study types other than economic evaluations was wider for the systematic review performed in this study compared to the one described in Chapter 2. From the model-based economic evaluation in Chapter 3 it was concluded that specialized psychotherapy for borderline personality has a high probability of cost-effectiveness. However, uncertainty remained regarding a societal cost perspective, longer time periods than 1 year, and cost-effectiveness relative to treatment-as-usual (TAU).

The advice as formulated in Chapter 2 to include all relevant societal costs and to use QALYs as an outcome measure could contribute to the improvement or standardization of the methodological quality of future studies. Such standardization would then also facilitate a synthesis of the empirical evidence (e.g., using the method that is presented in Chapter 3 or other methods), which is based on societal instead of only health care costs, on absolute instead of relative cost reductions, and on QALYs as assessed directly instead of via the remapping of e.g., BDI scores.

Chapters 4 and 5 presented an economic evaluation on group schema therapy (GST) for borderline personality disorder that is embedded in an international, multicentre RCT. In this RCT two formats for GST (one consisting of only group psychotherapy and one consisting of a combination of individual and group psychotherapy) and treatment as usual (TAU) are compared. Once completed, the results will add to both the empirical evidence base that is described in Chapter 2, as well as the one that is described in Chapter 3. Moreover, they will shed light on the important question whether a group format of schema therapy (whether or not in combination with individual sessions) is cost-effective in comparison to the optimal treatments that patients with BPD currently receive in the Netherlands.

The design of the RCT in Chapters 4 and 5 is well in line with the advice as formulated in Chapter 2, while all relevant societal costs are included, and QALYs are used as an outcome measure. This also applies to the advice as formulated in Chapter 3, since the duration of the RCT is longer than one year and the cost-effectiveness of specialized

psychotherapy is compared relative to TAU.

For the cost-effectiveness and cost-utility analyses in Chapter 5, a number of 5,000 cost-effectiveness pairs were simulated and visualized using an approach that was inspired by the model-based economic evaluation described in Chapter 3 to produce results in a way similar to the conventional bootstrapping approach. The specific advantage of this approach is that cost-effectiveness planes and cost-effectiveness acceptability curves can be constructed using all available data, including cases with missing values, without a need for imputation procedures to allow the calculation of total costs per patient.

The methodological study described in Chapter 6 was presented as an appropriate solution for the analysis of longitudinal cost-effectiveness data that includes cases with missing values. It presented a method for analysis that efficiently makes use of all available data with a coherent set of Bayesian multilevel models to describe the development of net benefit over time. It was considered as a useful extension to the net benefit regression framework as described by Hoch, Briggs, and Willan (2002).

Note that, as a consequence of their disorder (e.g., inability or unwillingness to participate due to psychological crises), attrition is characteristically high in clinical trials with patients suffering from personality disorders (e.g., Chapter 5 and Bamelis et al., 2014). Also, relatively long time periods are required when the interventions under investigation consist of psychotherapy for this patient group. The reason for this is twofold. First, the therapy itself is often delivered over a period of several years. Second, the impact of the intervention on societal costs and quality of life is likely to extend the duration of treatment. Therefore, the approach as outlined in Chapter 6 is of particular interest to research on personality disorders. For example, it was noted in Chapter 2 that not all economic evaluation studies perform an intention-to-treat analysis. Bayesian multilevel models could facilitate this; since cases with missing values do not need to be excluded from analysis and no additional procedures for the imputation of missing values need to be performed when this method is used.

7.3 Methodological considerations

This section presents a critical reflection on the following: the methodology behind the systematic review in Chapter 2, methodological aspects regarding the assessment of societal costs in Chapters 4 and 5, and the methodological study in Chapter 6.

7.3.1 Systematic review in Chapter 2

At the initiative of the Dutch 'Kenniscentrum Persoonlijkheidsstoornissen' (in English: 'Knowledge Centre Personality Disorders'), the systematic review in Chapter 2 (which was translated in English for this thesis) was submitted in reply to an invitation for a short contribution to a theme issue of the Dutch journal 'Tijdschrift voor Psychiatrie' (in English: 'Journal of Psychiatry') on the societal costs and benefits of psychiatry. Secondary to the aim of providing an up-to-date overview of studies, a quality assessment of the included economic evaluations was performed in this study.

Six criteria were used for the quality assessment. These provided a way of effectively summarizing the methodological aspects that were considered as the most important determinants of the quality of an economic evaluation. A more extensive, technical discourse on the methodological quality was not considered fitting given the aim and target audience of the study (as well as the short format for publication). However, the method that was used for the quality assessment was not based on the consensus criteria as established by leading experts in the field of health economics (e.g., Husereau et al., 2013).

7.3.2 The assessment of societal costs in Chapters 4 and 5

In addition to an assessment of clinical effectiveness, the economic evaluation in Chapter 4 and 5 required an assessment of societal costs. A structured interview was used for this, which has certain advantages. The structured nature of the interview ensures that the same set of questions is posed to each participant, and one can provide specific instructions to research assistants for each (set of) question(s). For example, the instruction to record typical examples of events that are related to costs, as a way of providing information that could facilitate a correct interpretation of the data at the time of analysis. However, the use of a structured interview and its reliance on patient report also has certain disadvantages that are related to coverage, subjectivity, and intricacy.

7.3.2.1 Coverage

A structured interview needs to be used that consists of a number of questions that is sufficient to cover any anticipated cost item with possible relevance to the disorder. In case of a (potentially) multifaceted disorder such as BPD, this implies that many cost items need to be assessed. A specific item can then easily be overlooked. For example, at the design stage of the cost interview that is used in the RCT described in Chapters 4 and 5 a specific item for the use of sheltered living accommodations was not included, and therefore not reported on. Also, patients may have had other costs relevant to the

disorder that are not asked for in the cost interview, or for which the description did not match with existing items.

7.3.2.2 Subjectivity

By relying on a patient's report, the assessment of societal costs per definition becomes prone to subjective interpretations, forgetting of relevant information, errors when calculations and estimations are required, et cetera. This can lead to data entries that are sometimes false, incomplete, or ambiguous. As such, it may be considered as suboptimal in comparison to registry-based cost assessment methods. However, many different registries would need to be consulted for the assessment of societal costs, and for many relevant cost items no registry exists. Registries themselves may contain errors as well, and an interview would still be needed for those items that are not recorded in any registry, such as out-of-pocket costs. Moreover, the use of patient data for the purpose of scientific research is restricted by Dutch law.

Similar to how subjective interpretation can be an issue in the use of a structured interview, a patient's eagerness to participate in an interview can have effects on data collection as well. A patient who is very eager to participate might constructively report every instance of his or her unhealthy eating behaviour as 'binge eating' even when in reality the behaviour can hardly be regarded as such. Inversely, a patient who genuinely binge eats (e.g., as a form of the self-damaging behaviour that is a symptom of BPD) might be unwilling to report this during the interview.

The difficulties related to subjectivity in patient report and the interpretation thereof also apply to both the research assistant who conducts the interview, and the analyst of the once entered data (e.g., in the form of a cost calculation). Even if the interviews for cost assessment were performed by highly experienced clinical experts, it still would not warrant error-free data collection. One approach that may help to harmonize the interpretation of data is that the analyst provides the research assistant, who conducts the cost interview, a detailed explanation of the rationale behind the data collection, the interpretation of different cost items or categories, and the formats for entering the data. For the study in Chapters 4 and 5 such explanatory sessions were given by telephone to the research assistants. Furthermore, the analyst was available to provide additional guidance when needed (e.g., to discuss the interpretation of a specific cost item before the data was entered in final form).

7.3.2.3 Intricacy

With hindsight, of some specific cost items (e.g., a large list of candy items of unknown prices that is reportedly consumed while binge eating) it is questionable whether it is

worthwhile to devote a relatively large amount of research time on its analysis, since its total impact on the incremental costs is likely to be small, and the accuracy and relevance of its assessment is doubtful. Similar reasoning could apply to the predetermined differentiation between various health care services (e.g., different types of social services), that in the end were valued using the same unit cost price, although, in the case of international, multicentre studies caution is needed since differences between countries may exist regarding which cost items are valued using the same cost prices. Furthermore, information on cost items that are (to a reasonable extent) similar in nature, and which are valued by cost prices with similar orders of magnitude (e.g., the cost prices for visits to a specialized mental health care facility, a psychotherapist, or psychiatrist), could likely be covered by one question and valued by a single unit cost price. After all, the societal costs are assessed as estimates, not exact measurements. Hence, it might be sensible to evaluate beforehand whether a detailed disentangling of the use of various similar items with similar cost prices is worth the effort, given that the 'signal-to-noise' ratio of the interview is inherently limited to a considerable extent.

A final point of consideration regarding the cost interview is the use of open-ended questions or 'miscellaneous' cost categories, since they have the potential to generate data that includes irrelevant items, being reported due to insufficient delimitation of the question. For example, costs for dental services were often reported, although they were not included in the study in Chapters 4 and 5, while these are unlikely to have a significant effect on the incremental costs between treatments. Nevertheless, open-ended questions or 'miscellaneous' cost categories also have the potential to generate unanticipated, yet relevant, data. These may pose a dilemma to the analyst, since the relevance of those costs warrants their inclusion, although it remains uncertain whether these potential costs have been assessed in other patients as well (i.e., other patients were not specifically asked to report these same costs), which in fact is a valid reason not to include them. In the trial-based economic evaluation that is described in Chapter 4 and 5, any costs that are reported in this way, and which are deemed relevant, are also included in the analysis whenever possible.

7.3.3 The methodological study in Chapter 6

The method that is presented in Chapter 6 is considered as a useful extension of the net benefit regression framework. Unfortunately, in its current implementation it does not provide the ability to produce cost-effectiveness planes. Hitherto, these are still the most common way of visualizing the sampling uncertainty using simulated (i.e., bootstrapped) cost-effects pairs in the health economics literature. Depending on the specific needs or reporting standards that an analysis is required to meet, the method presented in Chapter

6 may therefore not (yet) be suitable for application. For example, to meet the reporting standards as formulated in the application for the grant that subsidized the (Dutch part of) the study described in Chapter 4, and for which the results of the economic evaluation are reported in Chapter 5, a simulation procedure was performed using multilevel estimates for costs and effects in order to visualize the sampling uncertainty in a way that is similar to that when bootstrapped cost-effects pairs are displayed. Notwithstanding, cost-effectiveness acceptability curves (i.e., which essentially visualize the same information as cost-effectiveness planes, but in a different way) can readily be produced using the method.

The possibility to directly estimate a 'probability of cost-effectiveness' using the Bayesian approach in this study was considered as an advantage given the difficulties in inferring the same from a frequentist null hypothesis test (e.g., O'Hagan & Luce, 2003). In addition, the net benefit regression framework offered a solution to the difficulties caused by the two-dimensional nature of cost-effectiveness data (e.g., Briggs et al., 2002). The advantages of the Bayesian approach as well as those of the net benefit regression framework are combined in Chapter 6 with the advantages of using multilevel models for the analysis of longitudinal data. In addition to an efficient handling of cases with missing values, this also includes the ability to account for autocorrelation by a subset of the models.

Another advantage of Bayesian statistics is its flexibility in terms of specifying models as (mathematical relations between) parameters with their corresponding statistical distributions to account for the uncertainty in their values. For example, this flexibility is convenient for accommodating the skewness that is typically found in the statistical distributions of cost data. On the flipside, a caveat lures in this flexibility, since it offers generous opportunity for the misspecification of statistical models. This, as well as the fact that the Bayesian approach to statistics is not a common general component in academic teaching programs, may contribute to a steep learning curve for the applied researcher who is interested in using the approach. However, the advantages of Bayesian statistics, in combination with advances in the computational power of personal computers and (open source) software development, contribute to a recent trend in the further dissemination of Bayesian methods in various scientific research fields and academic programs. The methodological study described in Chapter 6, including the series of appendices, could contribute to this development.

7.4 Future prospects and recommendations

Based on the methodological considerations that were addressed in the previous section, prospects and recommendations for future studies are presented below. These include considerations regarding methodological choices for future studies as well as additional methodological avenues that could be explored.

7.4.1 Quality assessment of economic evaluations

As a follow up to the systematic review that is described in Chapter 2, a systematic review could be performed that is specifically targeted at an audience of (applied) health economists. For this study a quality assessment could be performed that is based on the criteria such as those from the CHEERS checklist (Husereau et al., 2013), the CHEC-list (Evers et al., 2005) or the checklist by Drummond (Drummond et al., 2015). This would allow for the inclusion of more criteria in the quality assessment overall as well as a better way of appraising whether the quality of an economic evaluation is up to par with the current recommendations as formulated by experts in the field.

7.4.2 Evidence synthesis

Following the recommendations for future economic evaluations that were discussed in Chapter 2 (i.e., for future studies to take into account all relevant societal costs and to use QALYs as an outcome measure), the use of a societal perspective for costs and QALYs as an outcome measure for effectiveness would contribute to an empirical evidence base of economic evaluations that are better comparable. This could facilitate the application of more conventional methods for evidence synthesis, such as meta-analysis. A further recommendation that is discussed in Chapter 3, and which follows the design of the RCT described in Chapter 4 and 5, pertains to how studies in the future could compare specialized psychotherapy for BPD with treatment as usual. This could help to resolve the uncertainty in the results regarding this aspect. Furthermore, it could pave the way for a sensible application of more sophisticated methods for evidence synthesis, such as network meta-analyses (e.g., Dias et al., 2018). These types of analyses could offer the possibility to synthesize evidence with the aim to assess the relative cost-effectiveness of the different types of specialized psychotherapy for borderline personality disorder on the basis of studies that provide direct (i.e., head-to-head comparisons between different types of psychotherapy) or indirect (i.e., comparisons to TAU) evidence. As discussed in Chapter 3, for future studies it is recommended to use a longer time period for follow up. That would allow an assessment of the cost-effectiveness and budget impact of psy-

chotherapy at the long term, which also pertains to the synthesis of empirical evidence on this aspect. Furthermore, availability of the information that is needed to calculate the budget impact of scaling up specialized psychotherapy to Dutch national level could make a better assessment of this important aspect possible.

7.4.3 Transferability

A limitation that was noted in the discussion of the model-based economic evaluation that is described in Chapter 3 applies to the fact that it remains unknown to which extent the results that are obtained in one country are transferable to another. The complete dataset of the international, multicentre RCT that is described in Chapter 4 will offer an opportunity to investigate this important aspect by assessing and quantifying the variability between the various components of societal costs.

7.4.4 Model-based economic evaluation

The complete dataset of the international, multicentre RCT that is described in Chapter 4 will furthermore offer a multitude of interesting data modelling options. Since this dataset will be rich in information on various aspects of borderline personality disorder, modelling techniques could be devised that integrate these different types of information. For example, one could speculate about modelling the relation between BPD symptom severity and societal costs. It could also be tested to what extent specific BPD symptoms correlate with specific costs. BPD symptoms such as anger or emotional instability, but also general measures of work and social functioning could be expected to correlate with productivity losses. Similarly, symptoms regarding self-damage and self-injury could correlate with costs due to out-of-pocket expenses (e.g., binge-eating, drug or alcohol use, impulsive spending) or certain health care costs (e.g., visits to general practitioners, and accidents & emergency departments). It could perhaps be possible to devise a Markov model (Briggs & Sculpher, 1998) consisting of different 'disease states' that represent the different combinations of (specific) BPD symptom severity and corresponding costs. Discrete-event simulation techniques (Caro et al., 2010) may be used in a similar way, with the additional potential to take into account the relations between past and subsequent events, in order to represent the occurrence of specific clinical conditions or use of health care resources. A particularly attractive modelling approach could be 'discretely integrated condition event' (DICE) simulation (Caro, 2016), which offers the potential to model a disease based on 'conditions' that persist or may change in their level over time as well as 'events' that occur at specific points in time. The different outcome measures that are used in the RCT as described in Chapter 4 may provide in-

formation for different BPD conditions in such a context, which is then integrated with information on health care resource use that could be modelled as events.

7.4.5 Assessment of societal costs

In light of the trial-based economic evaluation that is described in Chapters 4 and 5, an important aspect that deserves further scrutiny is the potential for improvements in the assessment and analysis of societal costs in future studies. Emphasis should always be put on the careful anticipation of all possible relevant costs at the design stage of an interview that is to be used for a specific patient population, in order to make sure that no important costs are missed at the time of assessment. For example, for a subsequent study on interventions for BPD the cost interview in its current form could be complemented by an item for the use of sheltered living accommodations.

The difficulties regarding subjective interpretation in the use of a cost interview could be decreased (i.e., in addition to providing detailed instructions to research assistants and careful anticipation of relevant cost drivers), by having the number of questions on different cost items to correspond better with the number of different cost prices for items that are used for valuation. This will effectively turn some cost items (this could apply as well to costs in the same order of magnitude) into 'cost categories', which has the advantage of being easier to interpret overall and making the overlooking or, inversely, the double-scoring of items less likely.

Similar reasoning could be applied to the level of detail in the assessment of patient out-of-pocket costs. For example, instead of asking for a (calculation that includes a) complete list of all the items that were consumed while binge eating, one could record the frequency of binge eating episodes with costs in the sub € 10 range (e.g., 'snacks' such as bags of chips or chocolate bars, et cetera) and those that were more expensive (e.g., ordered meals). These items could then be valued using estimated reference prices such as € 5 and € 20 for instance.

7.4.6 Bayesian multilevel net benefit regression

The methodological study that is described in Chapter 6 offers much potential for follow-up research. In addition to developing the method further as a 'two-dimensional' approach using both costs and effectiveness as the dependent variables as mentioned in the discussion of that chapter, further testing of the method is useful. The method could be contrasted with other approaches for dealing with data that includes cases with missing values, such as imputation-based approaches or even the simulation-based approach that is used for the synthesis of multilevel estimates for costs and effects in Chapter 5. In terms

of its application, it could offer a useful approach to the analysis of any longitudinal cost-effectiveness dataset that includes cases with missing values or for which insight into the development of net benefit over time is needed.

7.5 Implications for policy and practice

In terms of implications for policy, the studies in this thesis have their merit in terms of providing a) an up-to-date overview of the empirical evidence base for the cost-effectiveness of psychotherapy for BPD (as well as other personality disorders), b) an assessment of the cost-effectiveness of specialized psychotherapy for BPD in general and for group schema therapy specifically, and c) methodological advances for the analysis and synthesis of data on the cost-effectiveness of psychotherapy. Most of the evidence from previous studies (i.e., Chapter 2) as well as the results from the model-based economic evaluation (i.e., Chapter 3) suggest that psychotherapy for BPD is cost-effective. The results of the economic evaluation on group schema therapy for BPD (i.e., Chapter 4 and 5) will reveal whether the same applies to this treatment as well. As such, the studies in this thesis are relevant to policy makers. For example, empirical evidence can be used to inform decisions regarding reimbursement. Since the latter is an important factor in making treatments available to patients, this is relevant for clinical practice as well. For the practice of scientific research, the studies in this thesis provide important starting points for follow-up investigation and further methodological advances in future studies.

7.6 Conclusion

The studies in this thesis were performed to advance the assessment and analysis of the cost-effectiveness of psychotherapy for BPD. The topic was investigated using a variety of research methods: a systematic review, a model-based economic evaluation, a trial-based economic evaluation, and a methodological study. Each study on its own provides a point of reference for important follow-up investigation in the future. Together, the studies contribute to an overall evidence base that should serve as the starting point for any consideration regarding scientific follow-up research into the economic aspects of interventions for BPD, but also the updating of treatment guidelines, as well as policy decisions regarding the reimbursement of psychotherapy for patients with BPD.

7.7 References

Bamelis, L. L., Evers, S. M., Spinhoven, P., & Arntz, A. (2014). Results of a multicenter randomized controlled trial of the clinical effectiveness of schema therapy for personality disorders. *Am J Psychiatry*, *171*(3), 305-22.

Briggs, A. H., O'Brien, B. J., & Blackhouse, G. (2002). Thinking outside the box: recent advances in the analysis and presentation of uncertainty in cost-effectiveness studies. *Annual review of public health*, *23*(1), 377–401.

Briggs, A. H., & Sculpher, M. (1998). An introduction to Markov modelling for economic evaluation. *Pharmacoeconomics*, *13*(4), 397–409.

Caro, J. J. (2016). Discretely integrated condition event (DICE) simulation for pharmacoeconomics. *Pharmacoeconomics*, *34*(7), 665–672.

Caro, J. J., Möller, J., & Getsios, D. (2010). Discrete event simulation: the preferred technique for health economic evaluations? *Value in health*, *13*(8), 1056–1060.

Dias, S., Ades, A., Welton, N. J., Jansen, J. P., & Sutton, A. J. (2018). *Network meta-analysis for decision-making*. John Wiley & Sons.

Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes*. Oxford university press.

Evers, S., Goossens, M., De Vet, H., Van Tulder, M., & Ament, A. (2005). Criteria list for assessment of methodological quality of economic evaluations: Consensus on health economic criteria. *International journal of technology assessment in health care*, *21*(2), 240–245.

Hoch, J. S., Briggs, A. H., & Willan, A. R. (2002). Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Econ*, *11*(5), 415-30.

Husereau, D., Drummond, M., Petrou, S., Carswell, C., Moher, D., Greenberg, D., … Loder, E. (2013). Consolidated health economic evaluation reporting standards (CHEERS) statement. *Cost Effectiveness and Resource Allocation*, *11*(1), 6.

O'Hagan, A., & Luce, B. R. (2003). *A primer on Bayesian statistics in health economics and outcomes research.*

SUMMARY

Summary

This thesis presents a collection of studies that were performed to advance the assessment and analysis of the cost-effectiveness of psychotherapy for borderline personality disorder: a systematic literature review (Chapter 2), a model-based economic evaluation (Chapter 3), the study protocol (Chapter 4) and preliminary results (Chapter 5) of a trial-based economic evaluation, and a methodological study (Chapter 6).

Chapter 1 provides a general introduction to this thesis and explains the background of the studies. The nature and consequences of borderline personality disorder (BPD), the current optimal treatments for BPD, as well as the rationale and methodology of economic evaluation studies are addressed.

Chapter 2 presents a systematic review on economic evaluations of psychotherapy for personality disorders (Wetzelaer et al., 2016) that aims to provide an up-to-date overview of the studies that were previously performed in this field. The general characteristics of those studies as well as the specific characteristics of the economic evaluations, including an assessment of six important criteria regarding quality, are discussed. Most of the included studies indicate that psychotherapy for personality disorders is cost-effective. Yet, it is also noted that most studies did not include all (potentially) relevant societal costs, and that several studies did not use QALYs as an outcome measure. To make sure that future studies take all relevant costs and benefits into account, a (wider) societal perspective is advised. Furthermore, we recommend that future studies use QALYs as an (additional) outcome measure to facilitate the use of their results in reimbursement decisions.

Chapter 3 presents a model-based economic evaluation of the four specialized psychotherapies known to be effective for borderline personality disorder (BPD) specifically: dialectical behaviour therapy (DBT), schema therapy (ST), mentalization-based treatment (MBT), and transference-focused psychotherapy (TFP) (Wetzelaer et al., 2017). The starting point is a systematic literature review of studies that investigated one of these four psychotherapies. Included are studies that report information on the changes in health care costs or resource use, and/or quality of life (i.e., assessed directly using the Euroqol-5D or using the Beck Depression Inventory, of which the scores are mapped to the Euroqol-5D). This study aims to present a method for the synthesis of the evidence using simulated, patient-level data. The results suggest that specialized outpatient psychotherapy for BPD is cost-effective and that further extension of its supply in the Netherlands would require an investment of nearly € 2.4 million per 1,000 additional patients.

Chapter 4 presents the study protocol of an international multicenter randomized controlled trial (RCT) on group schema therapy (GST) for BPD, which includes an economic evaluation (Wetzelaer et al., 2014). The clinical effectiveness and cost-effectiveness

of two different formats of GST, one that consists of only group psychotherapy and one that consists of a combination of group and individual psychotherapy, and treatment as usual (TAU) are compared. In addition to the general background of the study, including the previous research that was performed in this specific field, information is given on the interventions being studied, the in- and exclusion criteria for participation, the outcome measures used for the assessment of clinical effectiveness of costs, and the scheduling of assessments. Furthermore, an outline is presented of the methodology behind the data analyses to be performed once data collection is completed. Finally, the strengths and limitations of the study design are discussed.

Chapter 5 presents the methods and preliminary results, obtained in the Netherlands, of the economic evaluation on GST for BPD. The relative cost-effectiveness and cost-utility of GST (both formats pooled) versus TAU is estimated, as well as that of GST-A (group schema therapy only) versus GST-B (group and individual schema therapy combined) versus TAU. The methods used for the assessment as well as the analyses of costs and effectiveness are addressed in detail. A simulation is performed to provide the estimates and figures regarding the relative cost-effectiveness and cost-utility of the interventions. These are displayed in cost-effectiveness planes and cost-effectiveness acceptability curves. Due to their preliminary nature, the results are blinded as to condition.

Chapter 6 presents a methodological study that aims to put forward how Bayesian multilevel models can be used as an extension of the net benefit regression framework. The approach provides an efficient way of handling longitudinal cost-effectiveness data that includes cases with missing values. A coherent set of models for the development of net benefit over time is described, as well as their application to an empirical example using the data from a previous RCT on schema therapy for personality disorders (Bamelis et al., 2015). The best fitting model is used to estimate the relative cost-effectiveness of schema therapy, clarification-oriented psychotherapy and TAU. The results are presented in cost-effectiveness acceptability curves that display the probability of relative cost-effectiveness for each intervention. Furthermore, the results are displayed using variants of the best fitting model that assume gamma or lognormal distributed data. In addition, a series of appendices is provided that explains the code and commands that are required to specify and run the models, as well as the statistical details behind the approach. It is concluded that Bayesian multilevel models provide an efficient and flexible method for the analysis of longitudinal cost-effectiveness data. Therefore, it is considered as a useful extension of the net benefit regression framework.

Chapter 7 provides a general discussion of the studies that were performed for this thesis. The studies and their findings are integrated to illustrate their overall sum of parts. Methodological considerations are provided that follow from a critical reflection on the methods used to perform the studies. Future prospects and recommendations for follow-

up investigation, and the implications for policy and practice are discussed. Lastly, a general conclusion of this thesis is provided.

References

Bamelis, L. L., Arntz, A., Wetzelaer, P., Verdoorn, R., & Evers, S. M. (2015). Economic evaluation of schema therapy and clarification-oriented psychotherapy for personality disorders: a multicenter, randomized controlled trial. *J Clin Psychiatry*, *76*(11), e1432-40.

Wetzelaer, P., Farrell, J., Evers, S., Jacob, G. A., Lee, C. W., Brand, O., … Arntz, A. (2014). Design of an international multicentre RCT on group schema therapy for borderline personality disorder. *BMC Psychiatry*, *14*(1), 319.

Wetzelaer, P., Lokkerbol, J., Arntz, A., van Asselt, A., & Evers, S. (2016). [Cost-effectiveness of psychotherapy for personality disorders; a systematic review on economic evaluation studies]. *Tijdschr Psychiatr*, *58*(10), 717-727.

Wetzelaer, P., Lokkerbol, J., Arntz, A., van Asselt, T., Smit, F., & Evers, S. (2017). Cost-effectiveness and budget impact of specialized psychotherapy for borderline personality disorder: A synthesis of the evidence. *Journal of Mental Health Policy and Economics*, *20*(4), 177–190.

Valorisation addendum

Valorisation addendum

Epiloguous to this thesis, this chapter serves to facilitate an appreciation of its added value in terms of the specific knowledge and products that were created. Given a financial interpretation, valorisation may easily be interpreted as 'turning something into money'. Yet, in a broader sense, e.g., when it is applied to knowledge, its meaning also refers to other kinds of benefit to (members of) society. For example, other scientists, health professionals or policy makers may utilize the knowledge that was created.

First, the societal and economic relevance of the studies is highlighted below. Subsequently, the (potential) stakeholders who could benefit from the work are described. Next, the innovativeness of the work is described. Finally, this chapter describes the specific purpose and possible use of the knowledge and products that were created in this thesis.

Societal and economic relevance

Borderline personality disorder (BPD) is a common and severe mental disorder, with a prevalence in the Netherlands of 1.1% in the general population (ten Have et al., 2016). In other countries, similar BPD prevalence estimates range between 0.5 and 2.7% (Samuels, 2011). Several studies have indicated that the quality of life of patients with BPD is severely impaired (Feenstra et al., 2012; IsHak et al., 2013; Perseius et al., 2006; Soeteman, Verheul, & van Busschbach, 2008). A diagnosis of BPD is a large burden to bear for a patient, and often also for his or her family, friends, colleagues, care providers, and others in his or her environment.

To society as a whole, the burden of BPD is significant in economic terms as well. To a large extent this can be explained by an extensive use of health care services, including both inpatient and outpatient facilities (Bender et al., 2001; Coid et al., 2009; Feenstra et al., 2012; Soeteman, Hakkaart-van Roijen, et al., 2008). Adding these costs to the costs of productivity losses (van Asselt et al., 2007; Soeteman, Hakkaart-van Roijen, et al., 2008) as well as other costs, such as those related to informal care and out-of-pocket costs (van Asselt et al., 2007), the total societal costs for BPD are substantial.

In the current Dutch health care system, the costs for specialized psychotherapy for BPD are not always (fully) reimbursed. On the one hand, this is because of the high costs of these treatments. On the other hand, it is due to the fact that the availability of empirical evidence regarding the clinical effectiveness and cost-effectiveness of (specific forms or formats for the delivery of) specialized psychotherapy for BPD is still limited. The result is a reluctant policy towards the reimbursement of such treatments, often in favor of treatment options with a shorter duration.

Psychotherapeutic treatments for BPD are costly, because a high number of (individual) psychotherapy sessions is often needed for succesful treatment. However, the costs for psychotherapy could possibly be partly offset by reductions in the costs for the use of other health care services. It therefore is important to determine how the potential differences in clinical effectiveness of psychotherapy for BPD relate to the potential differences in costs. This applies both to the extension of the supply of specialized psychotherapy for BPD (e.g., see Chapter 3 of this thesis) as well as to new formats for the delivery of (a specific type of) specialized psychotherapy for BPD, such as group schema therapy (GST; e.g., see Chapters 4 and 5 of this thesis). When psychotherapy can be effectively delivered (either in part or in full) in a group format, that enables a more efficient use of resources in comparison to only individual psychotherapy.

Economic evaluations are preferably performed from a societal perspective, so that all relevant costs and benefits to society are taken into account. The economic relevance of a study on the cost-effectiveness of health interventions is inherent to its definition as a study on 'the value for money' of such interventions. Cost-effectiveness studies may be informative for policy making, since they provide important indications for an efficient allocation of scarce (health care) resources.

Stakeholders

Patients & therapists

Studies on the cost-effectiveness of specialized psychotherapy for patients with BPD can have important implications for clinical practice. They contribute to the overall empirical evidence base that can be used to inform decisions regarding the reimbursement of these lenghty and intensive, and therefore costly, treatments. A better availability of treatments that are both clinically effective as well as cost-effective is beneficial for patients and therapists, as well as for society as a whole.

The results of a synthesis of the available empirical evidence on the costs and effects of specialized psychotherapy for patients with BPD indicate that it can be considered as a cost-effective treatment from a health care perspective (Chapter 3). This suggests that further investment to extend its supply would provide good value for money.

GST has been specifically designed as a specialized psychotherapy for patients with BPD, includes important additonal therapeutic elements that are not present in an individual setting and enhances efficiency in the use of health care resources through the use of group sessions for delivery. The results of the international, multicentre RCT described in Chapter 4 will reveal whether GST can be considered as a treatment that is both clinically effective, as well as cost-effective. Cost-effectiveness is assessed from a societal perspective, also taking into account patient and family costs, as well as costs in

other sectors (e.g., productivity losses), in addition to health care costs. A follow-up time period of three years is used to take into account costs and benefits of treatment beyond those of treatment duration. This study will provide important empirical evidence that can be used in decisions regarding the reimbursement of the costs for GST as well as the updating of treatment guidelines.

Scientists

Other scientists may benefit from the work in this thesis, since it provides examples of studies, as well as important directions for future research.

The overview of economic evaluations on psychotherapy for personality disorders in Chapter 2 could be an important starting point for any investigator with interest in the topic. Moreover, it provides important directions for the improvement of future studies (i.e., to perform economic evaluations from a societal perspective, and to use quality-adjusted life years as an outcome measure for effectiveness).

Chapter 3 of this thesis presents a method for the synthesis of empirical evidence on the costs and effects of specialized psychotherapy for BPD, that could be applied to other treatments and disorders as well. Furthermore, it may form a basis for future studies to further test, compare, and advance methodological approaches to the synthesis of evidence for cost-effectiveness and budget impact analyses. A detailed explanation of the methodology is provided and a mathematical appendix is included.

The empirical study described in Chapters 4 and 5 of this thesis will contribute to the evidence base regarding the cost-effectiveness of GST for BPD.

The methodological study in Chapter 6 of this thesis presents an application of Bayesian multilevel models to longitudinal cost-effectiveness data. The method offers an efficient approach to the analysis of longitudinal datasets that include cases with missing values. It makes use of Bayesian statistics to produce probabilistic statements about the relative cost-effectiveness of treatments, and can be flexibly extended to account for gamma and lognormal distributed data. It includes a detailed explanation of the underlying rationale and it demonstrates how the method is applied to an empirical dataset. A statistical appendix to this chapter furthermore provides the mathematical descriptions of the models, including prior distributions, as well as a manual to provide step-by-step instructions, including command syntax and model code, for the interested applied researcher. It could be helpful for any researcher interested in learning how to apply the method, as well as the expert analyst interested in further scrutinizing it.

Policy makers

The studies performed in this thesis are relevant for policy makers who decide over the reimbursement of the costs of psychotherapy for patients with BPD. When the results of these studies are used to inform decisions regarding reimbursement, this is also relevant for society. It could contribute to a more efficient allocation of scarce health resources, as well as better availability of effective treatment options for patients with BPD.

The synthesis of the available empirical evidence in Chapter 3 of this thesis suggests that specialized psychotherapy has a high probability of cost-effectiveness. This is important information in considering the (full) reimbursement of the costs for those treatments. Chapter 3 also provides important indications for future research to resolve those aspects that are still uncertain. This includes how specialized psychotherapy compares to treatment-as-usual (TAU), as well as when time horizons are used that are longer than one year.

The economic evaluation described in Chapter 4 and 5 of this thesis will provide empirical evidence on the comparative cost-effectiveness of two different formats for GST and TAU. A time horizon of three years is used. Importantly, this study is performed from a societal perspective (i.e., in contrast to the study in Chapter 3, which is performed from a health care perspective). This study will thus reveal whether GST, which is designed specifically as a specialized psychotherapy with enhanced efficiency in the use of health care resources, has added value in relation to current clinical practice, while taking into account an appropriate time horizon that covers the full two-year duration of treatment as well as an additional year, as well as all (potentially) relevant societal costs.

Since the economic evaluation that is performed alongside the RCT therefore takes into account all potentially relevant costs and effects and uses a follow-up time period that extends beyond treatment duration, it will allow policy makers to focus on not only the costs of the intervention, in addition to other changes in health care costs, but also on other costs that are relevant to society. For example, an effective psychotherapy that leads to recovery from BPD can also have beneficial effects through restorations of productivity losses. Societal costs other than those for health care are not of primary interest to health insurers, or other governing institutes with a focus on budgets that are restricted to health care costs. One could therefore question whether it makes sense to leave it to them, when a decision is required regarding the reimbursement of the costs for an intervention which has (economic) benefits that extend into other sectors, governmental departments, or governing institutes. Instead, it could be argued that intersectoral policy making is required for interventions with intersectoral costs and benefits. The study described in Chapters 4 and 5 will provide important insights on this aspect.

Innovativeness

The studies in this thesis are innovative in various ways. Chapter 2 provides an overview of economic evaluation studies on psychotherapy for BPD that was hitherto not available. Chapter 3 presents a method for the synthesis of data on costs and effects using all the relevant empirical findings that are available in the scientific literature. It has the advantage of incorporating various types of data, and from studies with different designs. Although the method is applied to the cost-effectiveness and budget impact of specialized psychotherapy for BPD, it could also be applied to the cost-effectiveness and budget impact of other treatments and disorders. Chapters 4 and 5 present a study on the (clinical and) cost-effectiveness of GST, which is an innovative approach to the treatment of BPD. Chapter 6 presents an innovative methodological approach to the analysis of longitudinal cost-effectiveness data using Bayesian multilevel models as an extension of the net benefit regression framework. The method offers an efficient approach to the analysis of longitudinal datasets that include cases with missing values. Furthermore, the use of Bayesian statistics enables probabilistic statements about the relative cost-effectiveness of treatments, which the use of null hypothesis tests, the method of inference used in traditional (frequentist) statistics, cannot.

Knowledge and products

The studies in Chapter 2, 3, and 4 are published as articles in the Dutch Journal of Psychiatry ('Tijdschrift voor Psychiatrie'), the Journal for Mental Health Policy and Economics, and BMC Psychiatry (BMC = BioMedCentral), respectively (Wetzelaer et al., 2016, 2017, 2014). These include both national and international, as well as peer-reviewed journals. Chapter 5 is based on the preliminary results of an international, multicentre RCT on group schema therapy, and is restricted to data from the Dutch sites. Due to their preliminary nature, these results are not amenable for publication. Once completed, the results of the RCT will add to the empirical evidence base for both the clinical effectiveness and cost-effectiveness of GST.

Transparancy is important when presenting methodologies, such as in Chapters 3 and 6, since it facilites fellow researchers who are interested in applying it, as well as those who wish to critically review it, build further upon it, or compare it with alternative methodologies. It therefore contributes to further advance the assessment and analysis of cost-effectiveness, in the context of psychotherapy for BPD, but also for other disorders, treatments, and beyond.

The study in Chapter 3 includes a detailed explanation of the methodology, as well as a mathematical appendix. The study in Chapter 6 includes a statistical appendix with

the mathematical descriptions of the models (including prior distributions). In addition, a detailed manual is included that provides step-by step instructions, including command syntax and model code, for the researcher who is interested in applying the method to a dataset. This will facilitate to make the knowledge that was created accessible to analysts and applied researchers without specialist expertise in Bayesian statistics.

References

Bender, D. S., Dolan, R. T., Skodol, A. E., Sanislow, C. A., Dyck, I. R., McGlashan, T. H., … Gunderson, J. G. (2001). Treatment utilization by patients with personality disorders. *American Journal of psychiatry*, *158*(2), 295-302.

Coid, J., Yang, M., Bebbington, P., Moran, P., Brugha, T., Jenkins, R., … Ullrich, S. (2009). Borderline personality disorder: health service use and social functioning among a national household population. *Psychol Med*, *39*(10), 1721-31.

Feenstra, D. J., Hutsebaut, J., Laurenssen, E. M., Verheul, R., Busschbach, J. J., & Soeteman, D. I. (2012). The burden of disease among adolescents with personality pathology: quality of life and costs. *J Pers Disord*, *26*(4), 593-604.

IsHak, W. W., Elbau, I., Ismail, A., Delaloye, S., Ha, K., Bolotaulo, N. I., … Wang, C. (2013). Quality of life in borderline personality disorder. *Harv Rev Psychiatry*, *21*(3), 138-50.

Perseius, K. I., Andersson, E., Asberg, M., & Samuelsson, M. (2006). Health-related quality of life in women patients with borderline personality disorder. *Scand J Caring Sci*, *20*(3), 302-7.

Samuels, J. (2011). Personality disorders: epidemiology and public health issues. *Int Rev Psychiatry*, *23*(3), 223-33.

Soeteman, D. I., Hakkaart-van Roijen, L., Verheul, R., & van Busschbach, J. J. (2008). The economic burden of personality disorders in mental health care. *Journal of Clinical Psychiatry*, *69*(2), 259-265.

Soeteman, D. I., Verheul, R., & van Busschbach, J. J. (2008). The burden of disease in personality disorders: diagnosis-specific quality of life. *J Pers Disord*, *22*(3), 259-68.

ten Have, M., Verheul, R., Kaasenbrood, A., van Dorsselaer, S., Tuithof, M., Kleinjan, M., & de Graaf, R. (2016). Prevalence rates of borderline personality disorder symptoms: a study based on the Netherlands mental health survey and incidence study-2. *BMC psychiatry*, *16*(1), 249.

van Asselt, A. D., Dirksen, C. D., Arntz, A., & Severens, J. L. (2007). The cost of borderline personality disorder: societal cost of illness in BPD-patients. *Eur Psychiatry*, *22*(6), 354-61.

Wetzelaer, P., Farrell, J., Evers, S., Jacob, G. A., Lee, C. W., Brand, O., … Arntz, A. (2014). Design of an international multicentre RCT on group schema therapy for borderline personality disorder. *BMC Psychiatry*, *14*(1), 319.

Wetzelaer, P., Lokkerbol, J., Arntz, A., van Asselt, A., & Evers, S. (2016). [Cost-effectiveness of psychotherapy for personality disorders; a systematic review on economic evaluation studies]. *Tijdschr Psychiatr*, *58*(10), 717-727.

Wetzelaer, P., Lokkerbol, J., Arntz, A., van Asselt, T., Smit, F., & Evers, S. (2017). Cost-effectiveness and budget impact of specialized psychotherapy for borderline personality disorder: A synthesis of the evidence. *Journal of Mental Health Policy and Economics*, *20*(4), 177–190.

Appendix 1

Sample size calculation for the multicentre RCT on GST for BPD in Chapter 4

Author of this appendix: Gerard van Breukelen

This appendix was first published as online supplement to Wetzelaer et al., 2014.

Design

Multicentre trial in 5 countries, 14 centres, a minimum of 2 cohorts of 16 patients per centre, randomization to group schema therapy (GST) versus treatment as usual (TAU) per cohort per centre. There are two formats for GST, GST-A and GST-B. Per centre one cohort will get GST-A or TAU, and the other cohort will get GST-B or TAU. Order of treatment will be balanced between centres; in half of the centres, cohort 1 will get GST-A or TAU and cohort 2 will get GST-B or TAU, and in the other half of the centres cohort 1 will get GST-B or TAU and cohort 2 will get GST-A or TAU.

Hypotheses

1. The outcome mean under group schema therapy (GST-A or GST-B) will be higher than under TAU.
2. The outcome mean under GST-A will differ from the outcome mean under GST-B.

Analysis

Data will be analyzed with mixed (multilevel) regression to take nesting of patients within centres into account, and adjusting for country, cohort and the difference between treatment A and B. For quantitative outcomes, the following mixed linear model will be applied for the outcome of patient i in centre j:

$$Y_{ij} = b_{0j} + b_{1j}cohort_{ij} + b_{2j}treatA_{ij} + b_{3j}treatB_{ij} + e_{ij}, \quad (1)$$

where cohort, treatA and treatB are dummy indicator variables for cohort (1 = cohort 2, 0 = cohort 1), GST-A (1 = GST-A, 0 = GST-B or TAU) and GST-B (1= GST-B, 0 = GST-A or TAU), respectively.

The regression weights are the sum of a fixed average weight β and a random centre effect μ, thereby allowing for a centre main effect, centre by cohort interaction, and centre by treatment interaction, respectively. The four random effects are allowed to covary. The last term, e_{ij}, is the

residual representing a random patient effect plus measurement error. The fixed model part will be extended with country effects using dummy coding, with cohort by treatment interaction, and with relevant covariates to increase power and to test for any hypothesized treatment by covariate interactions.

The model assumes one outcome measurement per patient. Repeated measures will be aggregated into a powerful summary measure following the methods in (Frison & Pocock, 1997) and (Senn et al., 2000). In case of a substantial percentage of missing values, this method will be replaced with three-level mixed regression analysis, adding time of measurement as third level below the patient level, and choosing as model for the treatment by time interaction the same model that underlies the choice of summary measure, i.e., linear divergence between treatment arms over time, allowing for nonlinear trend within each arm.

The model treats centre as a random effect. If the number of centres is too small for stable estimation of centre effects after adjusting for country, then centre will be included as fixed effect. This gives a smaller sample size than the present calculation, at the price of restricting all inferences to the centres in this trial.

Sample size calculation for a quantitative outcome and hypothesis 1

Since country, centre and cohort are orthogonal to both treatment indicators due to the design chosen, their fixed and random effects can be ignored in treatment effect estimation, giving the following contrasts of interest for hypothesis 1:

$$\frac{\mu_A + \mu_B}{2} - \frac{\mu_{TA} + \mu_{TB}}{2} \neq 0$$

where μ_A and μ_B are the expected outcomes under treatments GST-A and GST-B, and μ_{TA} and μ_{TB} are the expected outcomes under the TAU control to GST-A condition and under the TAU control to GST-B condition respectively. This contrast can be estimated by using the sample means per centre and averaging these across centres, assuming a sample size of n = 8 per treatment condition per cohort per centre.

Using model 1, the variance (= squared standard error) of this contrast estimator can be shown to equal:

$$\frac{\sigma_2^2 + \sigma_3^2 + 2\sigma_{23} + \frac{4\sigma_e^2}{n}}{4K}, \tag{2}$$

where K = the number of centres, n = the nr of patients per centre per cohort per treatment arm (we assume n = 8), and the variances are the between-centre variance of the GST-A effect, the between-centre variance of the GST-B effect, the between-centre covariance of the two effects, and the within-centre between-patient outcome variance. Assuming the between-centre variances

of GST-A effects and GST-B effects to be equal, the worst-case scenario is when the two treatment effects correlate perfectly between centres, reducing (2) to $(\sigma_2^2 + \frac{\sigma_e^2}{n})/K$, which is the same expression as for a multicentre trial with only one experimental and one control treatment arm and a total of 4n patients per centre (Moerbeek et al., 2000, 2003), noting that they used -1/+1 instead of 0/1 treatment coding which makes the treatment effect estimator, and its standard error twice as small, and σ_2^2 four times as small, as presently, for technical details, see (van Breukelen, 2013). Without centre by treatment interaction (i.e., if $\sigma_2^2 = 0$), it follows from standard sample size formulae (e.g., (Kirkwood, 1988) that a total of 168 patients (= 5.25 centres) is sufficient to detect an effect size of $d = 0.50$ with 90% power using a two-tailed α of 5%. Assuming centre by treatment interaction such that σ_2^2 is 5% of the between-patient σ_e^2 (which gives a typical intraclass correlation value of almost 0.05), we need 236 patients or 8 centres. Including 13 centres of 32 patients each will then give sufficient power for an effect size $d = 0.40$.

Sample size adaptation to hypothesis 2

The contrast of interest is now $\mu_A - \mu_B$ and the data from TAU do not add useful information here. In particular, subtracting from $\mu_A - \mu_B$ the term $\mu_{TA} - \mu_{TB}$ to adjust for cohort effects is superfluous as model (1) already adjusts for cohort effects, and will increase the standard error of the effect estimator. Hypothesis 2 is tested by running model (1) without the TAU patients and dropping the treatB indicator such that GST-B is reference treatment against which GST-A is compared. The contrast of interest has variance

$$\frac{\sigma_2^2 + \frac{2\sigma_e^2}{n}}{K}, \tag{3}$$

where σ_2^2 will have a different value than in model (1) for hypothesis 1, as it now reflects between-centre variance in the outcome difference between GST-A and GST-B rather than between GST-A and TAU. Compared with the result for hypothesis 1, the between-patient variance σ_e^2 counts twice since each treatment arm (GST-A versus GST-B) has now n patients per centre, against 2n for hypothesis 1 (GST-A and GST-B pooled versus TAU). Ignoring intraclass correlation gives a sample size of 168 patients (= 10.5 centres assuming 16 patients on treatment GST-A or GST-B per centre) to detect an effect of medium size $d = 0.50$ with 90% power using a two-tailed α of 5%. Taking again an intraclass correlation of nearly 0.05, we need a sample size of 202 patients (= 13 centres).

In short, a total of 13 centres will give a 90% power to detect an effect of size $d = 0.40$ for hypothesis 1 (TAU versus GST), and an effect of size $d = 0.50$ for hypothesis 2 (GST-A versus GST-B).Taking into account 5% attrition we thus need a total of 14 centres.

References

Frison, L. J., & Pocock, S. J. (1997). Linearly divergent treatment effects in clinical trials with repeated measures: efficient analysis using summary statistics. *Statistics in Medicine*, *16*(24), 2855-2872.

Kirkwood, B. R. (1988). *Essentials of medical statistics.* Oxford: Blackwell Scientific Publications.

Moerbeek, M., van Breukelen, G. J. P., & Berger, M. P. F. (2000). Design issues for experiments in multilevel populations. *Journal of Educational and Behavioral Statistics*, *25*(3), 271-284.

Moerbeek, M., van Breukelen, G. J. P., & Berger, M. P. F. (2003). A comparison between traditional methods and multilevel regression for the analysis of multicenter intervention studies. *Journal of clinical epidemiology*, *56*(4), 341–50.

Senn, S., Stevens, L., & Chaturvedi, N. (2000). Repeated measures in clinical trials: simple strategies for analysis using summary measures. *Statistics in Medicine*, *19*(6), 861-877.

van Breukelen, G. J. P. (2013). Optimal experimental design with nesting of persons in organizations. *Zeitschrift für Psychologie*.

Wetzelaer, P., Farrell, J., Evers, S., Jacob, G. A., Lee, C. W., Brand, O., … Arntz, A. (2014). Design of an international multicentre RCT on group schema therapy for borderline personality disorder. *BMC Psychiatry*, *14*(1), 319.

Appendix 2: Appendix Bayesian multilevel net benefit regression for longitudinal cost-effectiveness data

Appendix

Appendix

Bayesian Multilevel Net Benefit Regression for Longitudinal Cost-Effectiveness Data

Joran Jongerling
Pim Wetzelaer

Contents

1 Appendix A: Bayesian Multilevel Models

1.1 First Level Models

1.1.1 The First-Order Autoregressive Model

The first order autoregressive (AR(1)) model (Hamilton, 1994; Chatfield, 2003) describes a situation where there is no systematic change (no increase or decrease) over time, and where an individual's scores simply fluctuate around an individual mean. An AR(1) process can be expressed as consisting of two parts: a mean score μ, which represents an individual's trait score (i.e. his/her long-run tendency, equilibrium, or long-term preferred state) and a temporal deviation from this mean, which we denote as ζ_t, that is

$$y_t = \mu + \zeta_t. \tag{1}$$

The temporal deviations (or states) may be characterized by autocorrelation and can be modeled with the AR(1) model

$$\zeta_t = \phi\zeta_{t-1} + \varepsilon_t. \tag{2}$$

Where ϕ is the AR-parameter used to regress the current state on the previous one, and ε_t is the unpredictable part, also referred to as the innovation, residual, or random shock. It is assumed that ϕ lies between -1 and 1 to ensure stationarity (that is, a situation in which the mean and variance of the process do not change over time, see Hamilton, 1994, p. 53 and Chatfield, 2003, pg. 12).[1] Furthermore, it is assumed that the innovations are independent and normally distributed with a mean equal to 0 and variance σ^2_ε.[2]

The first order autoregressive (AR(1)) model as introduced by Hamilton (1994) and Chatfield (2003) can be extended to a multilevel model by allowing individuals to differ with respect to the model parameters μ and ϕ. This

[1] Note that the prior distribution for the AR-parameter is not restricted between -1 and 1 in the model code that is proved in the manual. However, it is monitored whether the estimates for this parameter stay within the acceptable limits that ensure stationarity of the model.

[2] Note that in the AR(1) model code the error variance (i.e. the proportion of the variance that cannot be predicted from the previous time point) for the first time point (i.e. tau1) differs from that of subsequent time points (i.e. tau2). This error variance cannot be estimated for the first time point, since no information is available from before that time point.

results in,

$$y_{it} = \mu_i + \zeta_{it}. \tag{3}$$

Where the temporal deviations can again be written as,

$$\zeta_{it} = \phi_i \zeta_{t-1} + \varepsilon_{it}, \tag{4}$$

with the model parameters being subject to the same asumption as mentioned above. Note that this multilevel extension of the AR(1) model is closely related to the simplex model as introduced by Jöreskog (1971, 1979). However, whereas the simplex model assumes that the model parameters are fixed across individuals (i.e. $\mu_i = \mu, \forall i$, and $\phi_i = \phi, \forall i$), the multilevel AR(1) model does not. It is important to realize however, that reliably estimating (inter-individual differences in) individual AR-parameters likely requires around 20, and preferably 50, repeated assessments for each participant (Jongerling, Laurenceau, & Hamaker, 2015). If fewer repeated assessments are available, as is the case in the current study, a multilevel AR(1) model in which only the mean is random might be preferred, in which case the amount of autocorrelation should be fixed across individuals (i.e. $\phi_i = \phi$ in Equation 4). This is exactly what was done for the analyses in the current study. In addition, we also assume that the amount of autocorrelation is constant over time in this study, even though the AR-parameter could theoretically be allowed to vary over time (e.g. the amount of autocorrelation between time points 1 and 2 could be modeled as different from the amount of autocorrelation between time points 2 and 3).

The AR-parameter ϕ_i in Equation 4 reflects the degree to which previous scores or states carry over into current scores or states. Suppose we have a number of repeated assessments of NB for an individual. If the AR-parameter is close to zero, this implies that there is little or no carry over from the level of NB on the previous assessment to the current level of NB. In contrast, when the AR-parameter is close to 1, this implies that an increased amount of NB at the previous assessment is likely to persist into the current assessment (and subsequent assessments), while decreased levels also tend to persist into subsequent assessments.

The innovation ϵ_{it} represents the part of the process that cannot be predicted based on previous scores or states. Thus, it can be thought of as the collection of all unobserved (or omitted) factors that influence the process under investigation.

In the AR(1) model, the growth curve for the average NB (i.e. the curve that shows the model predicted average NB at each of the assessments) is simply given by

$$\mu_t = \mu \qquad \text{for} \quad t = 1, \ldots, T, \tag{5}$$

where T is the total number of time points, and μ is the average NB across all individuals.

In the manual, we describe the R-code for this model and provide the model files for the Bayesian analysis with the JAGS-program (Plummer, 2003, 2012).

1.1.2 Latent Growth Model

The LGC model (Meredith & Tisak, 1990) is used to model linear change over time, and can be written as follows,

$$y_{it} = \alpha_i + \beta_i (t-1) + \varepsilon_{it}, \tag{6}$$

where α_i is an individual's intercept that represents his/her score at time point $t = 0$, β_i is the (linear) slope parameter which quantifies the amount of change between successive assessments, and ε_{it} is a random error term with zero mean and variance σ^2_ε. The residuals, ε_{it}, are assumed uncorrelated across individuals and time, and uncorrelated with the random intercept and random slope.

Note, that the slope parameter β_i is multiplied by the current assessment minus 1 $(t-1)$. This is done so the factor loading of the first assessment is equal to 0, making the first assessment our reference point, and an individual's intercept (α_i) equal to his or her expected NB score at this first assessment. A positive value for slope parameter β_i subsequently means that the scores get higher from each assessment to the next, while a negative value implies that the NB decreases over time.

For this model, the growth curve for the average NB is,

$$\mu_t = \mu_\alpha + (t-1)\mu_\beta \quad \text{for } t = 1, \ldots, T. \tag{7}$$

where μ_α and μ_β represent the mean intercept and mean slope across individuals respectively. Note that this equation shows that multiplying the slope by $t-1$ indeed makes the first assessment equal to the intercept. After all, filling in Equation 7 for $t = 1$ results in,

$$\mu_1 = \mu_\alpha + 0 * \mu_\beta, \tag{8}$$
$$\mu_1 = \mu_\alpha. \tag{9}$$

In the manual, we describe the R-code for this model and provide the model files for the Bayesian analysis with the JAGS-program (Plummer, 2003, 2012).

1.1.3 Quadratic Growth Model

The quadratic growth curve (QGC) model is very similar to the LGC model. The only difference is that the QGC model contains one additional parameter, the quadratic term (η_i), that allows for the modeling of quadratic growth over time. As a result, this type of model is better suited for modeling change that increases or decreases in rate over time. For example, if the increase (or decrease) in someone's NB is large at first, but then levels off, the QGC model can describe this type of change accurately, while the LGC, which assumes that the rate of change is the same across all assessments can not. The QGC model can even model change that is increasing at first and decreasing later on (or vice versa).

The QGC model can be written as follows,

$$y_{it} = \alpha_i + \beta_i(t-1) + \eta_i(t-1)^2 + \varepsilon_{it}, \tag{10}$$

where α_i and β_i are again the individual intercept and linear slope, η_i is an individual's quadratic term, and ε_{it} is again a random error term with zero mean and variance σ^2_ε, that is assumed to be uncorrelated across individuals and time, and uncorrelated with the other model parameters.

Note that the slope parameter β_i is again multiplied by the current assessment minus 1 $(t-1)$, and that the quadratic term is multiplied by $(t-1)^2$. Like with the LGC model, this is done to make the first assessment our reference point, and an individual's intercept (α_i) equal to his or her expected NB at this first assessment.

The growth curve for average NB of the QGC is equal to,

$$\mu_t = \mu_\alpha + (t-1)\mu_\beta + (t-1)^2\mu_\eta \quad \text{for } t = 1, \ldots, T. \tag{11}$$

where μ_α and μ_β are again the mean intercept and mean slope over individuals, while μ_η is the mean quadratic effect.

In the manual, we describe the R-code for this model and provide the model files for the Bayesian analysis with the JAGS-program (Plummer, 2003, 2012).

1.1.4 ALT model

The ALT model (Curran & Bollen, 2001;Bollen & Curran, 2004) is a combination of a LGC and an AR(1) model, and can be described as a LGC model with AR relations between the observed variables (Bollen & Curran, 2004). We included this model in the study because systematic change in NB over time is likely, but we also think that an individual's current amount of NB will depend on his/her previous amount. While the ALT model can model both these processes simultaneously, having both systematic change and autocorrelation between assessments introduces a difficulty.

This follows from the expression for the ALT model, which can be written as,

$$y_{it} = \alpha_i + \beta_i(t-1) + \phi y_{i,t-1} + \varepsilon_{it}, \tag{12}$$

where α_i is a constant, β_i is a change parameter, ϕ is the AR parameter, and ε_{it} is a random error term that is subject to the same assumptions as under the AR(1) and LGC model. What Equation 12 shows is that (unless all $\phi = 0$) there is a recursion in the ALT model. That is, y_{it} is a function of α_i, β_i, and $y_{i,t-1}$, but this last term is a function of α_i, β_i and $y_{i,t-2}$, and so on. Due to this recursion, we cannot use the ALT model unless we find a satisfactory way to incorporate all previous (unobserved) observations into the model. If the AR-parameter is restricted to be the same across the entire range of assessments and smaller than 1 in absolute value (i.e. $|\phi| < 1$), as it is in this study, then all previous assessments can be incorporated into the ALT model by putting nonlinear constraints on α_i (i.e. $(1-\phi)^{-1}$) and β_i (i.e. $\phi(1-\phi)^{-2}$) at the first assessment (Bollen & Curran, 2004) to get,

$$y_{i1} = \alpha_i(1-\phi)^{-1} - \beta_i\phi(1-\phi)^{-2} + z_{i1}, \text{ where} \tag{13}$$
$$z_{i1} = \varepsilon_{i1} + \phi\varepsilon_{i0} + \phi^2\varepsilon_{i,-1} + \cdots + = \sum_{j=0}^{\infty} \phi^j \varepsilon_{i,1-j}, \tag{14}$$

Assuming that the residual variation is equal across time, the term z_{i1} represents an infinite weighted sum of all the unobserved, previous residuals, which

thus solves for the recursion in the ALT model. In addition, this term has a zero mean and variance σ_z^2 (which is a function of σ_ε^2 and ϕ)(Hamaker, 2005). After Equations 13 and 14 are used to model the data at the first assessment, Equation 12 can be used for the subsequent assessments ($t = 2, 3, ..., T$). Although these constraints for α_i and β_i at $t = 1$ solve the recursion problem they make the parameters of the ALT model hard to interpret. Specifically, with these constraints, the intercept and slope of the ALT model are equal to (Hamaker, 2005),

$$\delta_i = \alpha_i(1-\phi)^{-1} - \beta_i\phi(1-\phi)^{-2}, \text{ and} \tag{15a}$$
$$\gamma_i = \beta_i(1-\phi)^{-1}, \tag{15b}$$

and not simply to α_i and β_i. For this variation of the ALT model, the growth curve for the average NB can be written as,

$$\begin{aligned}\mu_t &= \mu_\delta + (t-1)\mu_\gamma \\ &= \mu_\alpha(1-\phi)^{-1} - \mu_\beta\phi(1-\phi)^{-2} + (t-1)\mu_\beta(1-\phi)^{-1} \quad \text{for } t = 1, \ldots, T.\end{aligned} \tag{16}$$

Because of this difficulty with the interpretation of the parameters of the ALT model after using the constraints, we also consider another solution to the recursion problem.

This second solution involves rewriting the ALT model as a LGC model with autocorrelated disturbances, such as proposed by Chi and Reinsel (1989). This can be done because the ALT model and a LGC model with autocorrelated errors are algebraically equivalent when the AR-parameter is invariant over time (Hamaker, 2005), as it is in this study. The advantage of this second solution for the recursion problem, is that it separates the LGC, or trend, part of the model from the AR part, which keeps the interpretation of the model parameters straightforward. Specifically, a LGC model with autocorrelated errors can be written as,

$$y_{it} = \alpha_i + \beta_i(t-1) + \phi(y_{i,t-1} - (\alpha_i + \beta_i(t-2))) + \varepsilon_{it}. \tag{17}$$

where α_i and β_i are again the individual intercept and linear slope, ϕ is the autoregressive parameter which is subject to the same constraints as under the AR(1) model, and ε_{it} is again a random error. Note that β_i is

again multiplied by $(t-1)$ to make the first assessment our reference point and an individual's intercept (α_i) equal to his/her expected NB at this first assessment. For $t = 1$, the LGC model with autocorrelated errors can be written as,

$$y_{i1} = \alpha_i + \varepsilon_{i1}, \tag{18}$$

while for the subsequent occasions the model is,

$$y_{it} = \alpha_i + \beta_i(t-1) + \phi(y_{i,t-1} - (\alpha_i + \beta_i(t-2))) + \varepsilon_{it}. \tag{19}$$

Both error terms in these equations, ε_{i1} and ε_{it}, have a zero mean while the variance of ε_{i1}, $\sigma^2_{\varepsilon_{i1}}$, is equal to,

$$\sigma^2_{\varepsilon_{i1}} = \frac{\sigma^2_{\varepsilon_i}}{1-\phi^2}, \tag{20}$$

where $\sigma^2_{\varepsilon_i}$ is the variance of ε_{it}.

When writing the ALT model as an LGC model with autocorrelated disturbances, the intercept and the slope of the model are simply equal to α_i and β_i, while the parameter ϕ quantifies the amount of autocorrelation between successive assessment errors. The growth curve of average NB for this second variation is equal to,

$$\mu_t = \mu_\alpha + (t-1)\mu_\beta + \phi(\mu_{t-1} - (\mu_\alpha + (t-2)\mu_\beta)) \quad \text{for } t = 1, \ldots, T. \tag{21}$$

In the manual, we describe the R-code for this model and provide the model files for the Bayesian analysis with the JAGS-program (Plummer, 2003, 2012).

1.2 Second Level Model

On the second level inter-individual differences in the parameters of the AR(1), LGC, QGC, and ALT model are modeled (Jöreskog, 1971; Jöreskog, 1971; Meredith & Tisak, 1990; Bollen & Curran, 2004; Curran & Bollen, 2001). In this study we assume that all model parameters are normally distributed, making the second level expressions for models with only one random parameter equal to,

$$\mu_i \sim \mathcal{N}(\mu_\mu, \sigma^2_\mu), \tag{22}$$
$$\alpha_i \sim \mathcal{N}(\mu_\alpha, \sigma^2_\alpha), \tag{23}$$
$$\beta_i \sim \mathcal{N}(\mu_\beta, \sigma^2_\beta), \quad \text{and} \tag{24}$$
$$\eta_i \sim \mathcal{N}(\mu_\eta, \sigma^2_\eta). \tag{25}$$

When more than one parameter is random, the inter-individual differences in these parameters might be related. To account for this, the random parameter must be allowed to correlate with each other by giving them a joint distribution. Since we assume that our parameters are normally distributed, we will model the inter-individual differences in models with more than one random parameter using a multivariate normal (MVN) distribution. For a model with three random parameters, θ_1, θ_2, and θ_3, the MVN distribution can be written as,

$$\begin{bmatrix} \theta_{1i} \\ \theta_{2i} \\ \theta_{3i} \end{bmatrix} \sim MVN \left(\begin{bmatrix} \mu_{\theta_1} \\ \mu_{\theta_2} \\ \mu_{\theta_3} \end{bmatrix}, \begin{bmatrix} \tau^2_{\theta_1} & & \\ \tau_{\theta_1\theta_2} & \tau^2_{\theta_2} & \\ \tau_{\theta_1\theta_3} & \tau_{\theta_2\theta_3} & \tau^2_{\theta_3} \end{bmatrix} \right), \tag{26}$$

where $\tau^2_{\theta_1}$, $\tau^2_{\theta_2}$, and $\tau^2_{\theta_3}$ are the inter-individual variances in the model parameters; $\tau_{\theta_1\theta_2}$, $\tau_{\theta_1\theta_3}$, and $\tau_{\theta_2\theta_3}$ are the covariances between the random parameters; and μ_{θ_1}, μ_{θ_2}, and μ_{θ_3} are the expected values of θ_{1i}, θ_{2i}, and θ_{3i} respectively. For a LGC model (Meredith & Tisak, 1990), with a random intercept α_i and a random slope β_i, the level two model would be,

$$\begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} \sim MVN \left(\begin{bmatrix} \mu_\alpha \\ \mu_\beta \end{bmatrix}, \begin{bmatrix} \tau^2_\alpha & \\ \tau_{\alpha\beta} & \tau^2_\beta \end{bmatrix} \right). \tag{27}$$

When we want to add predictors on the second level of our models this is simply done by writing the means of the normal distributions in Equations 22 through 25 as linear equations containing these predictors. For example, assume patients receive three different treatments (like they did in this study) and we want to add treatment type as a predictor for inter-individual differences in the slope parameter of a LGC model β. This merely involves substituting the mean slope parameter μ_β from Equation 24 with,

$$\mu_{\beta i} = \gamma_{00} + \gamma_{10} * D1_i + \gamma_{20} * D2_i, \tag{28}$$

where $D1$ and $D2$ are two dummy variables used to identify which treatment individual i received, γ_{00} is the mean slope for the reference group, γ_{10} is the regression coefficient for *D1* which quantifies the difference in mean slope between the reference group and the group belonging to a score of 1 on *D1*, and γ_{20} is the regression coefficient for *D2* which quantifies the difference in mean slope between the reference group and the group belonging to a score of 1 on *D2*. For the other parameters in Equations 22 through 25 a similar approach can be used.

1.3 Bayesian Estimation of the Multilevel Models

In this study we use the JAGS program (Plummer, 2003, 2012) for the Bayesian estimation of our multilevel models. JAGS is a free software package and can be used by itself or in combination with R (R Core Team, 2014) using the R2jags package (Yu-Sung & Masanao, 2014) to call JAGS from R. The actual steps and code necessary to use JAGS to analyze data with the multilevel models discussed above are provided in a manual.

Here, we will focus a bit more on Bayesian estimation itself. In Bayesian estimation, several steps are required before a model can be estimated by a program (e.g. JAGS). While a thorough discussion of Bayesian statistics is beyond the scope of this article (interested readers are referred to Gelman, Carlin, Stern, and Rubin (2004), Hamaker and Klugkist (2011), and Hoijtink (2009)), there is one feature of Bayesian analysis that needs to be discussed here (albeit briefly): the prior distribution. In Bayesian statistics, researchers need to specify prior distributions for all model parameters, where these prior distributions represent a researcher's prior beliefs or knowledge about these parameters by assigning probabilities to their different possible values. These prior distributions are then combined with the distribution of the data using Bayes theorem in the following way,

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{f(y)}, \tag{29}$$

where y represents the sample data, and θ represents a model parameter we want to estimate (e.g. the AR-parameter ϕ of the AR(1) model, or the slope parameter β of the LGC model). In addition, $f(\theta|y)$ is the posterior distribution of parameter θ that represents the combined information about this parameter from both the prior and the data, $f(y|\theta)$ is the distribution of the data (y) conditional on parameter θ, $f(\theta)$ is the prior distribution for

parameter θ, and $f(y)$ is the marginal distribution of the data. Posterior distributions of parameters of interest are subsequently used for model estimation. That is, the mean, median, or mode of a posterior distribution can be used as the point estimate of a parameter, while the standard deviation of the posterior distribution can be seen as a measure of the sample variability of this estimate (analogous to the standard error in standard maximum likelihood estimation). If one has little or no prior knowledge, *vague (or diffuse)* priors can be used, which are characterized by assigning low and (approximately) equal probabilities to a very large range of possible values of a parameter. The results obtained with such priors depend almost exclusively on the data, and are therefore often close to Maximum Likelihood estimates.

For *fixed* parameters, that is, parameters that are not variance terms and that do not vary across individuals, normal distributions with 0 mean and (very) large variances are often used as vague priors, because the large variances spread the normal distribution out over a very large range of values, with each value in the range getting a very small and approximately equal probability. In this study, fixed effects will be assigned normal priors with 0 means and variances equal to 10^{12}, that is,

$$\pi \sim \mathcal{N}(0, 10^{12}), \tag{30}$$

where π represents the fixed parameter(s) of a model. For example, in the QGC model the priors for the mean intercept, mean linear change parameter, and mean quadratic change parameter are given by,

$$\mu_\alpha \sim \mathcal{N}(0, 10^{12}), \tag{31}$$
$$\mu_\beta \sim \mathcal{N}(0, 10^{12}), \tag{32}$$

and,

$$\mu_\eta \sim \mathcal{N}(0, 10^{12}). \tag{33}$$

For uncorrelated variance terms (e.g. the innovation variance σ^2_ε of a AR(1) model), the inverse gamma distribution is often used in Bayesian statistics, because this distribution does not allow values smaller than 0. The inverse gamma (IG) distribution is made vague in this study by setting

the two arguments of the distribution (the shape and the rate parameter) to the same value. In the current study we used inverse gamma distributions for residual variances and for the inter-individual variance of models with only one random parameter.[3] In these priors the shape and rate parameter were equal to 0.000001, so that,

$$\psi \sim IG(0.000001, 0.000001), \tag{34}$$

where ψ represents all uncorrelated variance parameters of a model. For an AR(1) model with a random mean μ, but fixed AR-parameter ϕ, the priors for the variance terms are,

$$\sigma^2_\varepsilon \sim IG(0.000001, 0.000001), \tag{35}$$

and,

$$\tau^2_\mu \sim IG(0.000001, 0.000001). \tag{36}$$

Finally, in models where more than one parameter is allowed to vary across individuals, it is possible that these random parameter are correlated with each other. To account for this, the variance terms in models with more than one random parameter are usually not assigned separate inverse gamma distributions, but are given a joint prior distribution, the Inverse Wishart (IW). This IW distribution is a prior for the entire covariance matrix of a model, and for a model with three random parameters θ_1, θ_2, and θ_3, it can be written as,

$$\begin{bmatrix} \tau^2_{\theta_1} & & \\ \tau_{\theta_1\theta_2} & \tau^2_{\theta_2} & \\ \tau_{\theta_1\theta_3} & \tau_{\theta_2\theta_3} & \tau^2_{\theta_3} \end{bmatrix} \sim IW(R, df), \tag{37}$$

where the left side of Equation 37 represents the covariance matrix of the model, R is a scale matrix that positions the distribution in multivariate space, and df are the degrees of freedom of the distribution which determine

[3] The use of gamma distributions for the prior distributions of hierarchical precision parameters is contentious (Gelman, 2006). Therefore, additional analyses to investigate sensitivity to prior distributions are advised.

how informative it is. For an AR(1) model with both a random mean μ and a random AR-parameter ϕ, the IW prior would be,

$$\begin{bmatrix} \tau_\mu^2 & \\ \tau_{\mu\phi} & \tau_\phi^2 \end{bmatrix} \sim IW(R, df). \tag{38}$$

Usually, an identity matrix is used for scale matrix R, but depending on the range of scores in the data this may not always be appropriate. In this study we will use data-based variance estimates on the diagonal of scale matrix R. To make the IW a vague prior, the df's need to be set to the number of random effects in the model, which would be 3 in the first example, and 2 for the AR(1) model with random mean and AR-parameter.

In the following sections we will specify the priors used for the different multilevel models fitted to the empirical data in this paper.

Based on preliminary convergence checks, the number of iterations for the Bayesian estimation procedures was set to 100,000 with a burn-in of 5,000.

1.3.1 Priors of the AR(1) Model with Fixed Mean and AR-parameter

For the AR(1) model with fixed mean and AR-parameter, both the mean and the AR-parameter are fixed effects. For both these parameters, normal distributions with 0 means and variances equal to 10^{12} were therefore chosen as priors. That is,

$$\mu \sim \mathcal{N}(0, 10^{12}) \tag{39}$$
$$\phi \sim \mathcal{N}(0, 10^{12}). \tag{40}$$

For the residual variance we specified an inverse gamma distribution as a prior, to get,

$$\sigma_\varepsilon^2 \sim IG(0.000001, 0.000001). \tag{41}$$

The R-code for this model is given in the manual.

1.3.2 Priors of the AR(1) Model with Random Mean and Fixed AR-parameter

In this variation of the AR(1) model, the (overall) mean and the AR parameter are fixed parameters, while the residual variance and the inter-individual variance in the mean are variance terms. Therefore, the prior distributions for the overall mean and AR-parameter, are equal to,

$$\mu_\mu \sim \mathcal{N}(0, 10^{12}) \tag{42}$$
$$\phi \sim \mathcal{N}(0, 10^{12}), \tag{43}$$

while the priors for the residual variance and the inter-individual variance in the mean are,

$$\sigma_\varepsilon^2 \sim IG(0.000001, 0.000001) \tag{44}$$
$$\sigma_\mu^2 \sim IG(0.000001, 0.000001). \tag{45}$$

The R-code for this model is given in the manual.

1.3.3 Priors of the AR(1) Model with Fixed Mean and Random AR-parameter

For the AR(1) model with fixed mean and random AR-parameter, both the mean and the (overall) AR-parameter are fixed effects. The priors for these parameters are,

$$\mu \sim \mathcal{N}(0, 10^{12}) \tag{46}$$
$$\mu_\phi \sim \mathcal{N}(0, 10^{12}). \tag{47}$$

For the residual variance and the inter-individual variance in the AR-parameter we specified inverse gamma distributions as priors, to get,

$$\sigma_\varepsilon^2 \sim IG(0.000001, 0.000001) \tag{48}$$
$$\tau_\phi^2 \sim IG(0.000001, 0.000001). \tag{49}$$

The R-code for this model is given in the manual..

1.3.4 Priors of the AR(1) Model with Random Mean and AR-parameter

In this variation of the AR(1) model, both the (overall) mean and the (overall) AR-parameter are fixed parameters. Therefore, for the overall mean and AR-parameter, the prior distributions are equal to,

$$\mu_\mu \sim \mathcal{N}(0, 10^{12}) \tag{50}$$
$$\mu_\phi \sim \mathcal{N}(0, 10^{12}), \tag{51}$$

The priors for the residual variance and the (possibly) related inter-individual variance in the mean and AR-parameter are,

$$\sigma^2_\varepsilon \sim IG(0.000001, 0.000001) \tag{52}$$
$$\begin{bmatrix} \tau^2_\mu & \\ \tau_{\mu\phi} & \tau^2_\phi \end{bmatrix} \sim IW(R, df), \tag{53}$$

where,

$$R = \begin{bmatrix} 1000 & \\ 0 & .01 \end{bmatrix} \text{ and,} \tag{54}$$
$$df = 2. \tag{55}$$

The R-code for this model is given in the manual.

1.3.5 Priors of the LGC Model with Fixed Intercept and Slope

In this model, the mean and slope are fixed effects, while the residual variance is a variance term, making the priors equal to,

$$\alpha \sim \mathcal{N}(0, 10^{12}) \tag{56}$$
$$\beta \sim \mathcal{N}(0, 10^{12}) \tag{57}$$
$$\sigma^2_\varepsilon \sim IG(0.000001, 0.000001) \tag{58}$$

The R-code for this model is given in the manual.

1.3.6 Priors of the LGC Model with Random Intercept and Fixed Slope

In this model, the overall intercept and slope are fixed effects, while the residual variance and the inter-individual variance in the intercept are variance terms, making the priors equal to,

$$\mu_\alpha \sim \mathcal{N}(0, 10^{12}) \tag{59}$$
$$\beta \sim \mathcal{N}(0, 10^{12}) \tag{60}$$
$$\sigma^2_\varepsilon \sim IG(0.000001, 0.000001) \tag{61}$$
$$\sigma^2_\alpha \sim IG(0.000001, 0.000001) \tag{62}$$

The R-code for this model is given in the manual.

1.3.7 Priors of the LGC Model with Fixed Intercept and Random Slope

In this model, the overall slope and intercept are fixed effects, while the residual variance and the inter-individual variance in the slope are variance terms, making the priors equal to,

$$\alpha \sim \mathcal{N}(0, 10^{12}) \tag{63}$$
$$\mu_\beta \sim \mathcal{N}(0, 10^{12}) \tag{64}$$
$$\sigma^2_\varepsilon \sim IG(0.000001, 0.000001) \tag{65}$$
$$\sigma^2_\beta \sim IG(0.000001, 0.000001) \tag{66}$$

The R-code for this model is given in the manual.

1.3.8 Priors of the LGC Model with Random Intercept and Slope

In this model, the overall slope and overall intercept are fixed effects, while the residual variance, the inter-individual variance in the intercept, and the inter-individual variance in the slope are variance terms. In addition, the random intercept and slope might be correlated, which means that we should use an Inverse Wishart prior for the random effects. The priors for this model are,

$$\mu_\alpha \sim \mathcal{N}(0, 10^{12}) \tag{67}$$
$$\mu_\beta \sim \mathcal{N}(0, 10^{12}) \tag{68}$$
$$\sigma^2_\varepsilon \sim IG(0.000001, 0.000001) \tag{69}$$
$$\begin{bmatrix} \tau^2_\alpha & \\ \tau_{\alpha\beta} & \tau^2_\beta \end{bmatrix} \sim IW(R, df). \tag{70}$$

where,

$$R = \begin{bmatrix} 1000 & \\ 0 & 500 \end{bmatrix} \text{ and,} \tag{71}$$
$$df = 2. \tag{72}$$

The R-code for this model is given in the manual.

1.3.9 Priors of the QGC Model with Fixed Intercept, Slope, and Quadratic Term

For this model the prior distributions are,

$$\alpha \sim \mathcal{N}(0, 10^{12}) \tag{73}$$
$$\beta \sim \mathcal{N}(0, 10^{12}) \tag{74}$$
$$\eta \sim \mathcal{N}(0, 10^{12}) \tag{75}$$
$$\sigma^2_\varepsilon \sim IG(0.000001, 0.000001) \tag{76}$$

The R-code for this model is given in the manual.

1.3.10 Priors of the QGC Model with a Random Intercept and Fixed Slope and Quadratic Term

For this model the prior distributions are,

$$\mu_\alpha \sim \mathcal{N}(0, 10^{12}) \tag{77}$$
$$\beta \sim \mathcal{N}(0, 10^{12}) \tag{78}$$
$$\eta \sim \mathcal{N}(0, 10^{12}) \tag{79}$$
$$\sigma^2_\varepsilon \sim IG(0.000001, 0.000001) \tag{80}$$
$$\sigma^2_\alpha \sim IG(0.000001, 0.000001) \tag{81}$$

The R-code for this model is given in the manual.

1.3.11 Priors of the QGC Model with Random Slope and Fixed Intercept and Quadratic Term

For this model the prior distributions are,

$$\alpha \sim \mathcal{N}(0, 10^{12}) \tag{82}$$
$$\mu_\beta \sim \mathcal{N}(0, 10^{12}) \tag{83}$$
$$\eta \sim \mathcal{N}(0, 10^{12}) \tag{84}$$
$$\sigma^2_\varepsilon \sim IG(0.000001, 0.000001) \tag{85}$$
$$\sigma^2_\beta \sim IG(0.000001, 0.000001) \tag{86}$$

The R-code for this model is given in the manual.

1.3.12 Priors of the QGC Model with a Random Quadratic Term and Fixed Intercept and Slope

For this model the prior distributions are,

$$\alpha \sim \mathcal{N}(0, 10^{12}) \tag{87}$$
$$\beta \sim \mathcal{N}(0, 10^{12}) \tag{88}$$
$$\mu_\eta \sim \mathcal{N}(0, 10^{12}) \tag{89}$$
$$\sigma^2_\varepsilon \sim IG(0.000001, 0.000001) \tag{90}$$
$$\sigma^2_\eta \sim IG(0.000001, 0.000001) \tag{91}$$

The R-code for this model is given in the manual.

1.3.13 Priors of the QGC Model with a Random Intercept and Slope, and a Fixed Quadratic Term

For this model, the priors are,

$$\mu_\alpha \sim \mathcal{N}(0, 10^{12}) \tag{92}$$
$$\mu_\beta \sim \mathcal{N}(0, 10^{12}) \tag{93}$$
$$\eta \sim \mathcal{N}(0, 10^{12}) \tag{94}$$
$$\sigma^2_\varepsilon \sim IG(0.000001, 0.000001) \tag{95}$$
$$\begin{bmatrix} \tau^2_\alpha & \\ \tau_{\alpha\beta} & \tau^2_\beta \end{bmatrix} \sim IW(R, df). \tag{96}$$

where,

$$R = \begin{bmatrix} 1000 & \\ 0 & 500 \end{bmatrix} \text{ and,} \tag{97}$$
$$df = 2. \tag{98}$$

The R-code for this model is given in the manual.

1.3.14 Priors of the QGC Model with a Random Intercept and Quadratic Term, and a Fixed Slope

For this model, the priors are,

$$\mu_\alpha \sim \mathcal{N}(0, 10^{12}) \tag{99}$$
$$\mu_\eta \sim \mathcal{N}(0, 10^{12}) \tag{100}$$
$$\beta \sim \mathcal{N}(0, 10^{12}) \tag{101}$$
$$\sigma^2_\varepsilon \sim IG(0.000001, 0.000001) \tag{102}$$
$$\begin{bmatrix} \tau^2_\alpha & \\ \tau_{\alpha\eta} & \tau^2_\eta \end{bmatrix} \sim IW(R, df). \tag{103}$$

where,

$$R = \begin{bmatrix} 1000 & \\ 0 & 500 \end{bmatrix} \text{ and,} \tag{104}$$
$$df = 2. \tag{105}$$

The R-code for this model is given in the manual.

1.3.15 Priors of the QGC Model with a Random Slope and Quadratic Term, and a Fixed Intercept

For this model, the priors are,

$$\mu_\beta \sim \mathcal{N}(0, 10^{12}) \tag{106}$$
$$\mu_\eta \sim \mathcal{N}(0, 10^{12}) \tag{107}$$
$$\alpha \sim \mathcal{N}(0, 10^{12}) \tag{108}$$
$$\sigma^2_\varepsilon \sim IG(0.000001, 0.000001) \tag{109}$$
$$\begin{bmatrix} \tau^2_\beta & \\ \tau_{\beta\eta} & \tau^2_\eta \end{bmatrix} \sim IW(R, df). \tag{110}$$

where,

$$R = \begin{bmatrix} 500 & \\ 0 & 500 \end{bmatrix} \text{ and,} \tag{111}$$
$$df = 2. \tag{112}$$

The R-code for this model is given in the manual.

1.3.16 Priors of the QGC Model with Random Intercept, Slope, and Quadratic Term

$$\mu_\alpha \sim \mathcal{N}(0, 10^{12}) \tag{113}$$
$$\mu_\beta \sim \mathcal{N}(0, 10^{12}) \tag{114}$$
$$\mu_\eta \sim \mathcal{N}(0, 10^{12}) \tag{115}$$
$$\sigma^2_\varepsilon \sim IG(0.000001, 0.000001) \tag{116}$$
$$\begin{bmatrix} \tau^2_\alpha & & \\ \tau_{\alpha\beta} & \tau^2_\beta & \\ \tau_{\alpha\eta} & \tau_{\beta\eta} & \tau^2_\eta \end{bmatrix} \sim IW(R, df), \tag{117}$$

where,

$$R = \begin{bmatrix} 1000 & & \\ 0 & 500 & \\ 0 & 0 & 100 \end{bmatrix} \text{ and,} \tag{118}$$

$$df = 3. \tag{119}$$

The R-code for this model is given in the manual.

1.3.17 Priors of the ALT Model with Fixed Constant and Change parameter

For this model, the priors are,

$$\alpha \sim \mathcal{N}(0, 10^{12}) \tag{120}$$
$$\beta \sim \mathcal{N}(0, 10^{12}) \tag{121}$$
$$\phi \sim \mathcal{N}(0, 10^{12}) \tag{122}$$
$$\sigma_\varepsilon^2 \sim IG(0.000001, 0.000001) \tag{123}$$

1.3.18 Priors of the ALT Model with Random Constant and Fixed Change parameter

For this model, the priors are,

$$\mu_\alpha \sim \mathcal{N}(0, 10^{12}) \tag{124}$$
$$\beta \sim \mathcal{N}(0, 10^{12}) \tag{125}$$
$$\phi \sim \mathcal{N}(0, 10^{12}) \tag{126}$$
$$\sigma_\varepsilon^2 \sim IG(0.000001, 0.000001) \tag{127}$$
$$\sigma_\alpha^2 \sim IG(0.000001, 0.000001) \tag{128}$$

1.3.19 Priors of the ALT Model with Fixed Constant and Random Change parameter

For this model, the priors are,

$$\alpha \sim \mathcal{N}(0, 10^{12}) \tag{129}$$
$$\mu_\beta \sim \mathcal{N}(0, 10^{12}) \tag{130}$$
$$\phi \sim \mathcal{N}(0, 10^{12}) \tag{131}$$
$$\sigma_\varepsilon^2 \sim IG(0.000001, 0.000001) \tag{132}$$
$$\sigma_\beta^2 \sim IG(0.000001, 0.000001) \tag{133}$$

1.3.20 Priors of the ALT Model with Random Constant and Change parameter

For this model, the priors are,

$$\mu_\alpha \sim \mathcal{N}(0, 10^{12}) \tag{134}$$
$$\mu_\beta \sim \mathcal{N}(0, 10^{12}) \tag{135}$$
$$\phi \sim \mathcal{N}(0, 10^{12}) \tag{136}$$
$$\sigma_\varepsilon^2 \sim IG(0.000001, 0.000001) \tag{137}$$
$$\begin{bmatrix} \tau_\alpha^2 & \\ \tau_{\alpha\beta} & \tau_\beta^2 \end{bmatrix} \sim IW(R, df). \tag{138}$$

where,

$$R = \begin{bmatrix} 1000 & \\ 0 & 500 \end{bmatrix} \text{ and,} \tag{139}$$

$$df = 2. \tag{140}$$

1.3.21 Priors of the ALT Model with Fixed Intercept and Slope

For this model, the priors are,

$$\alpha \sim \mathcal{N}(0, 10^{12}) \tag{141}$$
$$\beta \sim \mathcal{N}(0, 10^{12}) \tag{142}$$
$$\phi \sim \mathcal{N}(0, 10^{12}) \tag{143}$$
$$\sigma_\varepsilon^2 \sim IG(0.000001, 0.000001) \tag{144}$$

1.3.22 Priors of the ALT Model with Random Intercept and Fixed Slope parameter

For this model, the priors are,

$$\mu_\alpha \sim \mathcal{N}(0, 10^{12}) \tag{145}$$
$$\beta \sim \mathcal{N}(0, 10^{12}) \tag{146}$$
$$\phi \sim \mathcal{N}(0, 10^{12}) \tag{147}$$
$$\sigma_\varepsilon^2 \sim IG(0.000001, 0.000001) \tag{148}$$
$$\sigma_\alpha^2 \sim IG(0.000001, 0.000001) \tag{149}$$

1.3.23 Priors of the ALT Model with Fixed Intercept and Random Slope parameter

For this model, the priors are,

$$\alpha \sim \mathcal{N}(0, 10^{12}) \tag{150}$$
$$\mu_\beta \sim \mathcal{N}(0, 10^{12}) \tag{151}$$
$$\phi \sim \mathcal{N}(0, 10^{12}) \tag{152}$$
$$\sigma_\varepsilon^2 \sim IG(0.000001, 0.000001) \tag{153}$$
$$\sigma_\beta^2 \sim IG(0.000001, 0.000001) \tag{154}$$

1.3.24 Priors of the ALT Model with Random Intercept and Slope parameter

For this model, the priors are,

$$\mu_\alpha \sim \mathcal{N}(0, 10^{12}) \tag{155}$$
$$\mu_\beta \sim \mathcal{N}(0, 10^{12}) \tag{156}$$
$$\phi \sim \mathcal{N}(0, 10^{12}) \tag{157}$$
$$\sigma_\varepsilon^2 \sim IG(0.000001, 0.000001) \tag{158}$$
$$\begin{bmatrix} \tau_\alpha^2 & \\ \tau_{\alpha\beta} & \tau_\beta^2 \end{bmatrix} \sim IW(R, df). \tag{159}$$

where,

$$R = \begin{bmatrix} 1000 & \\ 0 & 500 \end{bmatrix} \text{ and,} \tag{160}$$
$$df = 2. \tag{161}$$

2 Appendix B: The DIC

Similar to other Information Criteria, like the AIC (Akaike, 1973) and BIC (Schwarz, 1978), the DIC (Spiegelhalter, Best, Carlin, & Linde, 2002) can be seen as consisting of two parts; one part that measures model (mis)fit, and a second part that quantifies the dimensionality, or complexity of a model. Specifically, the DIC can be written as,

$$\text{DIC} = -2\log f(y|\tilde{\theta}) + 2p_V. \tag{162}$$

where θ is a vector containing all model parameters (e.g. the AR-parameter ϕ, the mean parameter μ, and the residual variance σ_ε^2 of the AR(1) model with fixed mean and AR-parameter), $\tilde{\theta}$ is a Bayesian estimate of this vector, $-2\log f(y|\tilde{\theta})$ is a measure of model misfit (called the deviance) that represents the difference in fit between the model under consideration and a hypothetical "true" model that would fit the data perfectly, and p_V is an estimate of the complexity of the model expressed as the number of effective model parameters. The term $f(y|\tilde{\theta})$ that is part of the deviance is simply the density of the sample data given the Bayesian estimate of the parameter vector. Note that, since the deviance is a measure of misfit, lower values on the DIC imply a "better" model. In addition, Equation 162 shows that, like other model selection procedures, model selection based on the DIC is based on a trade-off between fit and complexity. If two models fit the data equally well (i.e. their deviances are equal), then the model with the lowest complexity will be selected.

Both the deviance and the number of effective model parameters are easily obtained through Gibbs-sampling (Gelman et al., 2004, p. 287-289). If we indicate the current iteration of the Gibbs-sampler with k (with $k = 1 \ldots K$) and the vector with the parameter estimates in iteration k as θ_k, the Bayesian estimate of the deviance can be written as,

$$-2\log f(y|\tilde{\theta}) = \frac{1}{K}\sum_{k=1}^{K}(-2\log f(y|\theta_k). \tag{163}$$

So the deviance can easily be obtained from the Gibbs-sampler by calculating the value of the deviance in each iteration of the Gibbs-sampler and subsequently calculating the mean of these values. Model complexity can be estimated in different ways using a Gibbs-sampler. Spiegelhalter originally

estimated the parameter p_D as a measure for the number of effective parameters in the model (Spiegelhalter et al., 2002). This value can be written as,

$$p_D = -2\log f(y|\tilde{\theta}) - (-2\log(y|\bar{\theta})) \text{ where,} \tag{164}$$

$$\bar{\theta} = \frac{1}{K}\sum_{k=1}^{K}(\theta_k). \tag{165}$$

So, Spiegelhalter subtracted the value of the deviance calculated at the mean values of the model parameters ($\bar{\theta}$) from the mean value of the deviance to get an estimate of model complexity. This method for calculating model complexity usually works fine, but is not invariant to reparametrization of the model, and can result in negative estimates for the effective number of parameters in certain situations. An alternative estimate for the effective number of parameters, suggested by Gelman et al. (2004) is estimate p_V,

$$p_V = \frac{1}{2K}\sum_{k=1}^{K}(-2\log f(y|\theta_k) - (-2\log f(y|\tilde{\theta})). \tag{166}$$

This estimate calculates the variance in the deviance values across the k iterations of the Gibbs-sampler and divides it by two to get the number of effective parameters. Estimate p_V solves the problems associated with p_D and is generally very robust and accurate. The only requisite is the use of vague priors. Given the advantages of p_V, that is the estimate we used in this study.

3 Appendix C: Model Extensions

In this paper we only applied our models to normally distributed nested data with two-levels. However, extending our models for the analysis of nested data with more than two levels and/or a non-normal distribution is also possible. In this section we will illustrate how our models can be extended to analyze data with three-levels, and how they can be extended to model lognormal or gamma distributed data.

3.1 Three-level Model

Clinical trials often take place in several (mental) health centres, and these centres can be seen as an additional level in our hierarchical data, that is, observations are nested within individuals, and these individuals are nested in the different centres. Extending the models above to include such a third (centre) level is straightforward, and only involves letting the individual level parameters (like the individual intercept α_i, lineair slope β_i, and quadratic slope η_i in the QGC model) come from distributions whose parameters depend on the centre c in which the individual was participating.

So, for three-level versions of AR(1), LGC, QGC, and ALT model, the parameters on the second level are distributed as,

$$\mu_i \sim \mathcal{N}(\mu_{\mu|c}, \sigma^2_{\mu|c}), \tag{167}$$
$$\alpha_i \sim \mathcal{N}(\mu_{\alpha|c}, \sigma^2_{\alpha|c}), \tag{168}$$
$$\phi_i \sim \mathcal{N}(\mu_{\phi|c}, \sigma^2_{\phi|c})\mathcal{I}_{|\phi_{ij}|<1\ \forall i,j}, \tag{169}$$
$$\beta_i \sim \mathcal{N}(\mu_{\beta|c}, \sigma^2_{\beta|c}), \quad \text{and} \tag{170}$$
$$\eta_i \sim \mathcal{N}(\mu_{\eta|c}, \sigma^2_{\eta|c}) \tag{171}$$

where $\mu_{\mu|c}$, $\mu_{\alpha|c}$, $\mu_{\phi|c}$, $\mu_{\beta|c}$, and $\mu_{\eta|c}$ are mean vectors containing $j = 1, ..., C$ separate mean parameter values. One for each of the C centres in the study. Similarly, $\sigma^2_{\mu|c}$, $\sigma^2_{\alpha|c}$, $\sigma^2_{\phi|c}$, $\sigma^2_{\beta|c}$, and $\sigma^2_{\eta|c}$ are vectors containing $j = 1, ..., C$ separate variance estimates. As was the case for the second level models presented in Appendix A (Equations 22 through 25), Equations 167 through 171 only apply when there is just one random parameter in a model (or when the random model parameters are not correlated). In case a model contains two or more random parameters, the rank order in one might depend on the

other(s). If this is the case, we need to account for it by giving the parameters a joint distribution. As with the two-level models in Appendix A we choose a multivariate normal (MVN) distributions as our joint parameter since we assume our parameters are normally distributed (Equation 26).

For the $j = 1, ..., C$ separate values in vectors $\mu_{\mu|c}$, $\mu_{\alpha|c}$, $\mu_{\phi|c}$, $\mu_{\beta|c}$, and $\mu_{\eta|c}$ we also assume normal distributions:

$$\mu_{\mu|c,\ j} \sim \mathcal{N}(0, 10^{12}), \forall j \in \mu_{\mu c}, \tag{172}$$
$$\mu_{\alpha|c,\ j} \sim \mathcal{N}(0, 10^{12}), \forall j \in \mu_{\alpha c}, \tag{173}$$
$$\mu_{\phi|c,\ j} \sim \mathcal{N}(0, 10^{12}), \mathcal{I}_{|\phi_{ij}|<1\ \forall i,j}, \forall j \in \mu_{\phi c}, \tag{174}$$
$$\mu_{\beta|c,\ j} \sim \mathcal{N}(0, 10^{12}), \forall j \in \mu_{\beta c}, \quad \text{and} \tag{175}$$
$$\mu_{\eta|c,\ j} \sim \mathcal{N}(0, 10^{12}), \forall j \in \mu_{\eta c}. \tag{176}$$

The parameters $\sigma^2_{\mu|c}$, $\sigma^2_{\alpha|c}$, $\sigma^2_{\phi|c}$, $\sigma^2_{\beta|c}$, and $\sigma^2_{\eta|c}$ contain $j = 1, ..., C$ separate variance terms. As was the case for the second-level variance terms, we assume that each of these follow an inverse-gamma distribution in models in which there is only one random parameter (or in which the random parameters are uncorrelated),

$$\sigma^2_{\mu|c,\ j} \sim \mathcal{IG}(0.000001, 0.000001), \forall j \in \mu_{\mu c}, \tag{177}$$
$$\sigma^2_{\alpha|c,\ j} \sim \mathcal{IG}(0.000001, 0.000001), \forall j \in \mu_{\alpha c}, \tag{178}$$
$$\sigma^2_{\phi|c,\ j} \sim \mathcal{IG}(0.000001, 0.000001), \forall j \in \mu_{\phi c}, \tag{179}$$
$$\sigma^2_{\beta|c,\ j} \sim \mathcal{IG}(0.000001, 0.000001), \forall j \in \mu_{\beta c}, \quad \text{and} \tag{180}$$
$$\sigma^2_{\eta|c,\ j} \sim \mathcal{IG}(0.000001, 0.000001), \forall j \in \mu_{\eta c}. \tag{181}$$

When more than one parameter is random at the same time, we need to account for possible covariances between these parameters. As was the case for the fixed effects (Equations 167 through 171) this is achieved by specifying a joint distribution for two or more variance terms, and as we did for the inter-individual variances on level 2, we choose an inverse Wishart distribution for this (Equation 37). This distribution is a prior for an entire variance-covariance matrix and can be seen as a multivariate version of the IG distribution.

For our best fitting model, the QGC model with random intercept and slope, a JAGS model file for running a three-level analysis is provided in the manual.

3.2 Lognormal and Gamma Distributed Data

In the above we assumed that the data were normally distributed. However, cost data in economic evaluations of health interventions are typically skewed. This is due to the fact that most patients in a trial usually incur relatively low costs and only a few patients incur high costs, for instance because of adverse events leading to hospitalizations. Therefore, it may be more reasonable to model such data using heavy right-tailed parametric distributions, such as lognormal or gamma distributions.

Fortunately, our code can also be applied to these other distributions.[4] If we want to apply a QGC model to lognormal data for example, one merely uses the logarithm of the data as the dependent value, making the equation for this model equal to,

$$\log(y_{it}) = \alpha_i + \beta_i(t-1) + \eta_i(t-1)^2 + \varepsilon_{it}, \tag{182}$$

For gamma distributed data, one has to change the specified distribution of the data in the model file that is read into the JAGS program. For the QGC model this would result in,

$$y_{it} \sim \text{Gamma}(\kappa, \nu_{it}) \quad \text{where,} \tag{183}$$

$$\nu_{it} = \kappa / \left(\alpha_i + \beta_i(t-1) + \eta_i(t-1)^2\right), \tag{184}$$

where the parameters κ and ν are the shape and rate parameters of the gamma distribution. Since the mean of a gamma distribution is equal to,

$$\frac{\mu_\kappa}{\mu_\nu}, \tag{185}$$

where μ_κ is the mean shape parameter and μ_ν is the mean rate parameter across individuals, Equation 184 makes sure that the mean trend of the gamma distributed data follows a quadratic growth curve. For the QGC model, the JAGS model file for Gamma distributed data is available in the manual. For the other models, the model files can be adapted for gamma distributed data by changing the specified distribution of y_{it}, and setting the rate parameter $\nu_{i,t}$ equal to the shape parameter κ divided by the appropriate mean trend.

[4] A change of scale, such as when performing a logarithmic transformation or gamma regression in combination with a log-link function, has important implications for the interpretation of parameters (including prior distributions).

4 Appendix D: Preliminary analysis

In a preliminary analysis we plotted the overall mean trajectories of net benefit over time for a range of different willingness-to-pay (WTP) values (Figure 1). We also plotted the mean trajectories of net benefit over time per condition (for a WTP value of €40.000; Figure 2), as well as the individual trajectories (Figure 3, Figure 4 and Figure 5). These plots are useful to guide the initial selection of a subset of models that are consistent with the results of these preliminary analyses as well as for the comparison with model-predicted trajectories as an additional check of model fit (i.e. complementary to the use of the DIC). For example, note how the slight curvature in the (overall) mean trajectories for the development of net benefit over time is consistent with a QGC model providing the best fit to these data.

Figure 1: Overall mean trajectories for willingness-to-pay (WTP) values of €0 to €80.000.

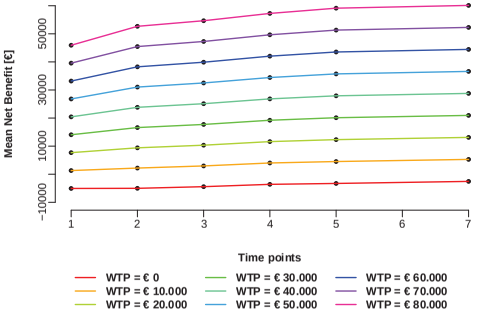


Figure 2: Mean trajectories per condition for a willingness-to-pay value of €40.000.

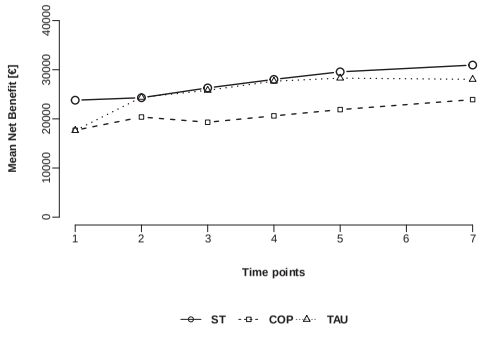


Figure 3: Individual trajectories in ST for a willingness-to-pay value of €40.000.

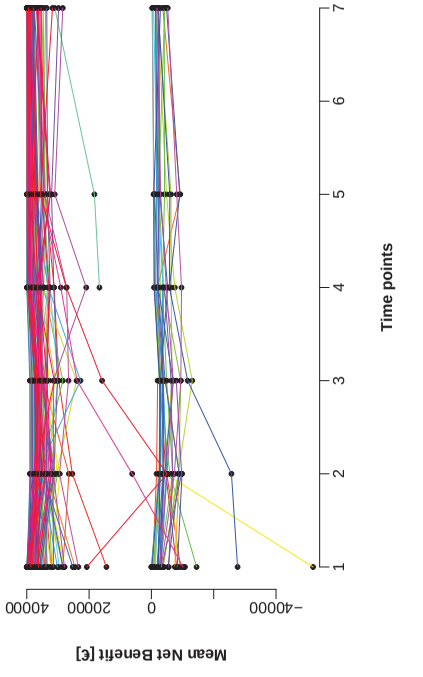


Figure 4: Individual trajectories in COP for a willingness-to-pay value of €40.000.

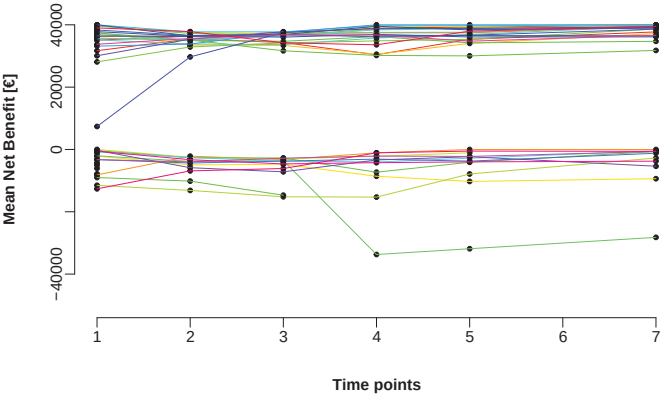
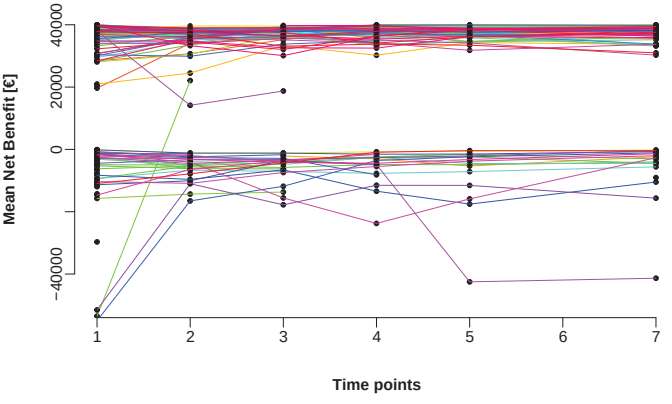


Figure 5: Individual trajectories in TAU for a willingness-to-pay value of €40.000.



References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. Petrov & F. Caski (Eds.), *Proceedings of the second international symposium on information theory.*

Bollen, K., & Curran, P. (2004). Autoregressive latent trajectories (ALT) model: A synthesis of two traditions. *Sociological Methods and Research*, *32*, 336-383.

Chatfield, C. (2003). *The analysis of time series: An introduction* (6th, Ed.). Washington, DC: Chapman & Hall/CRC.

Chi, E., & Reinsel, G. (1989). Models for longitudinal data with random effects and AR(1) errors. *Journal of the American Statistical Association*, *84*, 452-459.

Curran, P. J., & Bollen, K. A. (2001). The best of both worlds: Combining autoregressive and latent curve models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (p. 105-136). Washington, DC: American Psychological Association.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, *1*, 515 - 533.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2004). *Bayesian data analysis*. New York, NY: Chapman & Hall.

Hamaker, E. L. (2005). Conditions for the equivalence of the autoregressive latent trajectory model and a latent growth curve model with autoregressive disturbances. *Sociological Methods and Research*, *33*, 404-416.

Hamaker, E. L., & Klugkist, I. (2011). Bayesian estimation of multilevel models. In K. Roberts & J. Hox (Eds.), *Handbook of advanced multilevel analysis* (p. 137-161). New York: Taylor and Francis.

Hamilton, J. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.

Hoijtink, H. (2009). Bayesian data analysis. In A. Millsap R.E. & Maydeu-Olivares (Ed.), *The sage handbook of quantitative methods in psychology* (p. 423-443). London: SAGE.

Jongerling, J., Laurenceau, J. P., & Hamaker, E. L. (2015). A multilevel AR(1) model: Allowing for inter-individual differences in trait-scores, inertia, and innovation variance. *Multivariate Behavioral Research*, *50*(3), 334-349.

Jöreskog, K. G. (1971). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology*, *23*, 121-145.

Jöreskog, K. G. (1979). Statistical estimation of structural models in longitudinal-developmental investigations. In J. Nesselroade & P. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (p. 303-352). New York: Academic Press.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107-122.

Plummer, M. (2003). Jags: A program for analysis of Bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing (dsc 2003)*.

Plummer, M. (2012). Jags version 3.3.0 user manual [Computer software manual].

R Core Team. (2014). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461-464.

Spiegelhalter, D., Best, N., Carlin, B., & Linde, A. v. d. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *64(4)*, 583-639.

Yu-Sung, S., & Masanao, Y. (2014). R2jags: A package for running jags from r [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=R2jags` (R package version 0.04-01)

Manual

Manual

Bayesian Multilevel Net Benefit Regression for Longitudinal Cost-Effectiveness Data

Joran Jongerling
Pim Wetzelaer

Contents

1 Introduction

In the study by Wetzelaer, Jongerling, Arntz, and Evers (2017) we used several Bayesian multilevel models for the analysis of longitudinal cost-effectiveness data from (multi-centre) randomized controlled trials (RCTs). Specifically, we fit a First-Order Autoregressive (AR(1)) Model, a Latent Growth Curve (LGC) Model, a Quadratic Growth Curve (QGC) Model, and a Autoregressive Latent Trajectory (ALT) Model to the net benefit (NB) of several treatments for borderline personality disorder. In this short manual we illustrate how applied researchers can fit these models to their own data by describing 1) the required software, 2) the required format of the input files, 3) the actual model-code and where to find it, 4) the analysis of lognormal or gamma distributed data, and 5) the output generated by the code. Note that the aim of this manual is merely to aid applied researchers in *running* the code on their own data. As such, this manual will not go into details on the theory and math behind the estimation. Bayesian Statistics, for example, will not be extensively discussed. For this, and for a more in-depth explanation of the different models and their components, the interested reader is referred to the article by Wetzelaer et al. (2017).

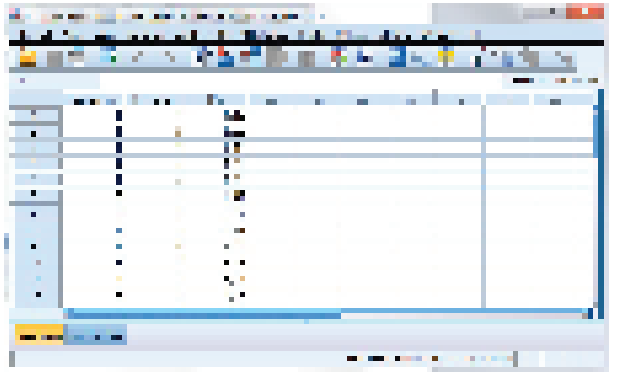
2 Required Software

To fit the Bayesian multilevel models, researchers need to install both the R statistical software (R Core Team, 2014) and the JAGS (Just Another Gibbs Sampler) program (Plummer, 2012). These programs are available from the websites `www.r-project.org/` and `mcmc-jags.sourceforge.net/`, respectively. The R program is needed to load and (if needed) manipulate researchers' data, set up the Bayesian analyses, and store the output, while the JAGS program is needed for the actual computations of the analyses using Markov Chain Monte Carlo (MCMC) simulation. The JAGS program can be called from inside R however, using the R2jags package (Yu-Sung & Masanao, 2014), so users do not have to switch between the two programs to analyze their data. The R2jags package can be installed from within the R-program by running the command `install.packages(R2jags)` and subsequently running `library(R2jags)` to load the package.

3 Format of the Input Files

Researchers can provide their nested data in both the so called 'long' and 'wide' formats. In the 'long' format, the repeated assessments of the dependent variable are all stored as scores on one variable, and each observation is entered on a separate line in the dataset. The individual and assessment to which each score belongs are subsequently indicated by two separate variables. Figure 1 shows part of an SPSS dataset that is structured in the 'long' format.

Figure 1: SPSS data in 'long' format.

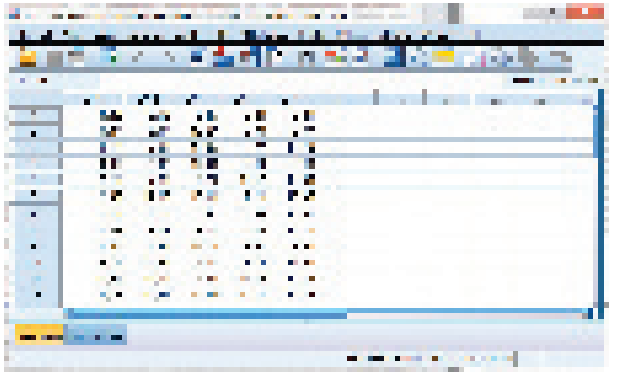


In these example data, the variable *Participant* is used to identify the different participants in the study, the variable *Occasion* identifies the different assessments, and the variable *DV* contains the individual scores on the dependent variable. The first five entries on the *DV* variable can then be identified as the repeated assessments for participant 1 on occasion 1 through 5. Similarly, the scores on *DV* entered on line 6 through 10 are the repeated assessments belonging to participant number two.

In the 'wide' format the repeated assessments of the dependent variable

are stored as scores on *separate* variables, and all the information of an individual is entered on a *single* line in the dataset. Figure 2 shows the same data as Figure 1 but structured according to the 'wide' format.

Figure 2: SPSS data in 'wide' format.



In these example data, variables *DV1* through *DV5* represent the repeated assessments of the DV on occasion 1 through 5. All the information of participant 1 is now presented on the first row of the dataset, while all scores of participant 2 are on the second row. Once the data is structured according to one of the two appropriate formats, the dataset must be saved as a .txt file, and can then be read into R.

4 Model-Code

4.1 Loading the Data

An R-package containing an overall function to run all the code for the analyses is currently still in the making, but until then, applied researchers can

run all the necessary analyses using the R-code available in Appendix A. In this section we present a short walk-through of this code.

To fit the Bayesian Multilevel Models to their data, applied researchers first need to read their datafile (structured using the instructions in the previous section) into R. This is done by the following lines of R-code.

```
Yrepeated <- read.csv("[Path To File]", sep = "", header = TRUE,
                      na.strings = "999999.00")
```

where the part that reads [Path To File] is replaced with the full pathname of the datafile (e.g. "C:/Study/Repeated/Datafile.txt"). Note that we are using 999999 as our missing values label, but this can be changed to any label that the researcher wants.

If the datafile was already in the 'wide' format (mentioned above and shown in Figure 2), the data is now correctly read in, and readers can skip the rest of this section and continue in section 4.2. However, if the data is in the 'long' format shown (mentioned above and shown in Figure 1), it needs to be restructured to the 'wide' format before it can be analyzed with the code presented in this manual. This is done with the following lines of code. Here, we assume that, next to the repeated assessments (located in the fourth column of the dataset), there are three more variables in the dataset; two dummy variables ($D1$ and $D2$) located in the second and third column of the dataset that indicate group membership, and one variable (REC) located in the fifth column that indicates whether a participant recovered or not. However, extending the code to include more additional variables and/or variables located in different columns is straightforward.

```
YWidth <- matrix(NA,nrow=(nrow(Yrepeated))/6,ncol=6)
D1 <- matrix(NA,nrow=(nrow(Yrepeated))/6,ncol=1)
D2 <- matrix(NA,nrow=(nrow(Yrepeated))/6,ncol=1)
Yrec <- matrix(NA,nrow=(nrow(Yrepeated))/6,ncol=6)
ind <- unique(Yrepeated[,1])

for (i in 1:(nrow(Yrepeated)/6)){
YWidth[i,] <- Yrepeated[Yrepeated[,1]==ind[i],4]
}

for (i in 1:(nrow(Yrepeated)/6)){
```



```
D1[i,] <- mean(Yrepeated[Yrepeated[,1]==ind[i],2])
}

for (i in 1:(nrow(Yrepeated)/6)){
D2[i,] <- mean(Yrepeated[Yrepeated[,1]==ind[i],3])
}

for (i in 1:(nrow(Yrepeated)/6)){
    Yrec[i,] <- Yrepeated[Yrepeated[,1]==ind[i],5]
}

Yrepeated <-  matrix(cbind(ind,D1,D2,YWidth,Yrec),nrow=
              length(ind))
```

The data is now ready to be sent to JAGS for analysis.

4.2 Analyzing the Data using JAGS

The JAGS program is specifically designed for Bayesian inference using MCMC simulation. Since JAGS uses Bayesian statistics it needs two additional pieces of information next to the data to be analyzed. First it needs a file describing the model to be estimated and the prior distributions for the model parameters, written in the BUGS language (D. J. Lunn, Thomas, Best, & Spiegelhalter, 2000). Second, it needs starting values for all the model parameters in order to start up the Gibbs-sampler (Gelman, Carlin, Stern, & Rubin, 2004, p. 287-289). The model files for our models are available in Appendix B, while a more in-depth discussion of the models is available in Wetzelaer, Jongerling, Arntz, and Evers (2017). The starting values for the model parameters are specified in the call to the R2jags package that is used to run JAGS from within R. Before any of the model can be fitted to the data we first need to load the R2jags package into R using the following code,

```
library(R2jags)
```

Next, the model specific code can be run.

4.2.1 AR(1) Model

One can fit four versions of the AR(1) model to the data; 1) an AR(1) model with a fixed mean and fixed AR-parameter, 2) an AR(1) model with a

random mean and fixed AR-parameter, 3) an AR(1) model with fixed mean and random AR-parameter, and 4) an AR(1) model in which both parameters are random. The code for the AR model with the fixed mean is,

```
initsFixedMuAR <- function(){
  list(mu=rnorm(1), AR=runif(1,0,1))
  list(mu=rnorm(1), AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]

FixedMuAR <- jags(data = list("DV","T","N"), initsFixedMuAR,
             model.file="[Path To File]", parameters.to.save=
             c("mu","var.y", "AR"), n.chains =2, n.iter = Itt,
             n.burnin=Bi,n.thin=1, DIC=TRUE).

FixedMuAR <- autojags(FixedMuAR, n.iter = Itt, n.thin=1,
 n.update= 10, DIC=T)

FixedMuAR.mcmc <- as.mcmc(FixedMuAR)
traceplot(FixedMuAR.mcmc)
autocorr.plot(FixedMuAR.mcmc)
```

The first lines of code generate two lists with starting values for the model parameters by taking a random draw from a standard normal distribution for the mean[1], and by taking a draw from a uniform distribution over the interval (-1,1) for the AR-parameter. These two lists are subsequently stored in the variable *initsFixedMuAR*.

The next three lines are used to specify the data that need to be sent to JAGS, where the part between square brackets again needs to be changed by the researcher to match his data. The important thing here is that for the

[1]To avoid the risk that the two chains may start too close together, it is advised to specify plausible and appropriate values for each chain separately instead of randomly generating starting values.

DV, the part between square brackets needs to specify those columns of the datamatrix created when the data was read into R that contain the repeated scores on the DV. The variables T and N simply contain the number of repeated assessments and the number of people in the sample, respectively.

The next line of code is the call to the JAGS program using the *jags* command form the R2jags package. The arguments for this call are as follows. First the sample data is defined in a list containing the names of the three variables that were just created for 1) the repeated assessments on the dependent variable, 2) the number of time points, and 3) the sample size. The second argument specifies the variable that contains the lists with starting values for the Gibbs-sampler (Gelman et al., 2004, p. 287-289). In the third argument, the model file to be used in the analysis is specified, and the part between the quotation marks should be replaced by the complete pathname for this file (e.g. C:/Study/Repeated/FixedARmodel-JAGS.txt). The specific model file needed for this model is given in section 10.1 through 10.4 of this manual. A zip-file containing JAGS model files for all models discussed in this manual are available upon request. The fourth argument specifies for which model parameters the researcher wants to get estimates. Here we specify that we only want estimates for the mean ("mu"), the total variance of the time series ("var.y"), and the AR-parameter. One can ask for additional model parameters (if there are any) in this argument, as well as for the scores on the dependent variable ("DV"). Asking for these scores will result in the output of the JAGS program containing the original dataset with any missing values imputed. The fifth argument specifies the desired number of MCMC chains to run on the data. Note that this number and the number of lists containing starting values should always match (so two chains need to be accompanied by two separate lists of starting values, as is the case here). We recommend specifying two or more separate MCMC chains so that the Gelman-Rubin statistic (Gelman et al., 2004, pg. 296-298), a statistic that gives some indication to whether the Gibbs-sampler converged properly, can be calculated by the JAGS program. The Gelman-Rubin statistic and model convergence will be mentioned in more detail in the section on the output of our model code. The next three arguments *n.iter*, *n.burnin*, and *n.thin* determine the behavior of the Gibbs-sampler. The first of these arguments specifies the total number of iterations of the Gibbs-sampler, while the second specifies the burn-in, that is, the number of iterations that will be used for the Gibbs-sampler to get to the appropriate posterior distribution of the saved model parameters. These burn-in iterations will be discarded and will

therefore not influence the final parameter estimates. Based on our experiences with the models presented here, we advise a value of 100.000 for *n.iter* and a value of 5.000 for *n.burnin*. If the Gibbs-sampler does not converge (see section 8 of the model), these numbers can be increased. The argument n.thin, specifies the thinning rate which indicated how many of the iterations of the Gibbs-sampler are kept for parameter estimation. Specifically, a thinning rate of 1 indicates that all iterations are used, while a thinning rate of 2 means that every other iterations is used. This argument is only of interest when one wants to discard intermediate iterations of the Gibbs-sampler because the correlation between successive iterations is too high. High autocorrelation decreases the efficiency of the posterior sample and accuracy of inferences (D. Lunn, Jackson, Best, Thomas, & D., 2013, Section 4.5). Usually, this argument can be set to 1. The final argument specifies whether the researcher want the Deviance Information Criterion (DIC) (Spiegelhalter, Best, Carlin, & Linde, 2002) value for the model, and can be equal to either "TRUE" or "FALSE". The DIC is a measure of model fit, and out of a set of models, the model with the lowest DIC value is the model that fits the sample data best.

The last three lines are again used to check for convergence of the Gibbs-sampler and to solve for any non-convergence issues. As mentioned above, this will be explained later in the *Output* section (section 8).

For the AR(1) model with the random mean and fixed AR-parameter the code is very similar,

```
initsRandomMuFixedAR <- function(){
  list(mean.mu=rnorm(1), AR=runif(1,0,1))
  list(mean.mu=rnorm(1), AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]

RandomMuFixedAR <- jags(data = list("DV","T","N"),
   initsRandomMuFixedAR,
   model.file="[Path To File]",
   parameters.to.save= c("mean.mu","var.y",
   "AR"), n.chains =2, n.iter = Itt,
```



```
   n.burnin=Bi, n.thin=1, DIC=T)


RandomMuFixedAR <- autojags(RandomMuFixedAR, n.iter = Itt,
   n.thin=1, n.update=10,DIC=T)

RandomMuFixedAR.mcmc <- as.mcmc(RandomMuFixedAR)
traceplot(RandomMuFixedAR.mcmc)
autocorr.plot(RandomMuFixedAR.mcmc)
```

The only difference is that we now have separate means for each individual in the dataset, so instead of specifying starting values for "mu" we now specify starting values for the average value of this mean parameter across individuals "mean.mu". In addition, we also want an estimate for "mean.mu", so we have to include this parameter in the *parameters.to.save* argument. Note that adding the parameter "mu" to the *parameters.to.save* argument will now result in JAGS providing estimates of all the individual means.

For the AR(1) model with the fixed mean and random AR-parameter the code is ,

```
initsFixedMuRandomAR <- function(){
  list(mu=rnorm(1), mean.AR=runif(1,0,1))
  list(mu=rnorm(1), mean.AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]

FixedMuRandomAR <- jags(data = list("DV","T","N"),
   initsFixedMuRandomAR,
   model.file="[Path To File]",
   parameters.to.save= c("mu","var.y",
   "mean.AR"), n.chains =2, n.iter = Itt,
   n.burnin=Bi, n.thin=1, DIC=T)
```

```
FixedMuRandomAR <- autojags(FixedMuRandomAR, n.iter = Itt,
   n.thin=1, n.update=10,DIC=T)

FixedMuRandomAR.mcmc <- as.mcmc(FixedMuRandomAR)
traceplot(FixedMuRandomAR.mcmc)
autocorr.plot(FixedMuRandomAR.mcmc)
```

The only difference with the code for the previous model is that we now have a random AR-parameter instead of a random mean. This means that we now have separate AR-parameters for each individual in the dataset, so instead of specifying starting values for "AR" we now specify starting values for the average value of the AR-parameter across individuals "mean.AR". In addition, we also want an estimate for "mean.AR", so we have to include this parameter in the *parameters.to.save* argument. Note that adding the parameter "AR" to the *parameters.to.save* argument will now result in JAGS providing estimates of all the individual AR-parameters.

Finally the code for the AR(1) model with both a random mean and a random AR-parameter is given by,

```
initsRandomMuAR <- function(){
  list(mean.mu=rnorm(1), mean.AR=runif(1,0,1))
  list(mean.mu=rnorm(1), mean.AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]

RandomMuAR <- jags(data = list("DV","T","N"), initsRandomMuAR,
              model.file="[Path To File]", parameters.to.save=
              c("mean.mu","var.y", "mean.AR"), n.chains =2,
              n.iter = Itt, n.burnin=Bi, n.thin=1, DIC=T)


RandomMuAR <- autojags(RandomMuAR, n.iter = Itt, n.thin=1,
  n.update= 10,DIC=T)
```



```
RandomMuAR.mcmc <- as.mcmc(RandomMuAR)
traceplot(RandomMuAR.mcmc)
autocorr.plot(RandomMuAR.mcmc)
```

As for the AR(1) model with a random mean we specify a starting value for the average value of this mean parameter across individuals, and like we did in the AR(1) model with a random AR-parameter we specify an initial value for the average value of the AR-parameter across individuals "mean.AR". In addition, we now want estimates for these two average parameter values, so we have to include both "mean.mu" and "mean.AR" in the *parameters.to.save* argument. Note that adding the parameter "mu" and/or "AR" to the *parameters.to.save* argument will now result in JAGS providing estimates of all the individual means and/or AR-parameters.

The code for the other models is very similar to the code for these two AR(1) models since the arguments for the call to JAGS will always be the same. Nevertheless we will shortly discuss the code for the LGC, QGC, and ALT model below, and we will highlight any differences between the codes of the different models.

4.2.2 LGC Model

For the LGC model we provide code for four different variations depending on which parameters are random. Specifically, we provide code for 1) a LGC model with a fixed intercept and a fixed slope, 2) a LGC model with a random intercept and a fixed slope, 3) a LGC model with a fixed intercept and a random slope, and 4) a LGC model with a random intercept and a random slope. The code for the first of the variations of the LGC, the LGC with a fixed intercept and a fixed slope, is as follows,

```
initsLGCFMS <- function(){
  list(alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100))

  list(alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100))

}
```

```
DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different
 Measurement Occasions]

LGCFMS <- jags(data = list("DV","T","N","Occasion"),
  initsLGCFMS, model.file="[Path To File]",
  parameters.to.save= c("alpha","var.y", "beta"),
  n.chains =2, n.iter = Itt, n.burnin=Bi, n.thin=1,
  DIC=T)

LGCFMS <- autojags(LGCFMS, n.iter = Itt, n.thin=50,
  n.update= 10, DIC=T)

LGCFMS.mcmc <- as.mcmc(LGCFMS)
traceplot(LGCFMS.mcmc)
autocorr.plot(LGCFMS.mcmc)
```

The structure of this code is again similar to that of the two AR(1) models mentioned above. The first lines of code specify two lists of starting values for the model parameters, where *alpha* represents the intercept of the LGC model and *beta* represents the slope, and stores these lists in a variable *initsLGCFMS*. The next four lines of code again specify the data that need to be sent to JAGS, where the part between square brackets again needs to be changed by the researcher to match his or her data. In addition to the variables also created for the AR(1) model the LGC model also requires the specification of an indicator variable for the different assessments *Occasion*. This *Occasion* variable is merely a list of integer values (starting at 1), with one value for each assessment. So, if there are 7 assessments, the *Occasion* variable is simply a vector containing the number 1 through 7.[2] The next line of code is again the call to the JAGS program, which has the same arguments as were discussed with the AR(1) models.

[2]Alternatively, the different assessments could be coded using the numbers 0 through 6. In that case, it is important to replace Occasion[t] - 1 with Occasion[t] in the model code.

When we allow either the intercept (*alpha*), the slope (*beta*), or both to be random across individuals, the changes to the code of the LGC model are similar to the changes made to the AR(1) code when we allow for intra-individual differences in the mean. Specifically, the parameter that is allowed to be random will be replaced by its average value across individuals in the list of starting values. In addition, parameter estimates for this average value across individuals will now be asked for in the *parameters.to.save* argument (although the individual estimates for the parameter can also be requested). So, for the LGC model with a random intercept, the code becomes,

```
initsLGCRMFS <- function(){
  list(mean.alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100))

  list(mean.alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

LGCRMFS <- jags(data = list("DV","T","N","Occasion"),
   initsLGCRMFS, model.file="[Path To File]",
   parameters.to.save= c("mean.alpha","var.y", "beta"),
   n.chains =2, n.iter = Itt, n.burnin=Bi, n.thin=1,
   DIC=T)


LGCRMFS <- autojags(LGCRMFS, n.iter = Itt, n.thin=1,
   n.update= 10, DIC=T)

LGCRMFS.mcmc <- as.mcmc(LGCRMFS)
traceplot(LGCRMFS.mcmc)
autocorr.plot(LGCRMFS.mcmc)
```

For the LGC model with a random slope the code becomes,

```
initsLGCFMRS <- function(){
  list(alpha=rnorm(1, mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100))

  list(alpha=rnorm(1, mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

LGCFMRS <- jags(data = list("DV","T","N","Occasion"),
   initsLGCFMRS, model.file="[Path To File]",
   parameters.to.save= c("alpha","var.y", "mean.beta"),
   n.chains =2, n.iter = Itt, n.burnin=Bi, n.thin=1,
   DIC=T)


LGCFMRS <- autojags(LGCFMRS, n.iter = Itt, n.thin=1, n.update= 10,
           DIC=T)

LGCFMRS.mcmc <- as.mcmc(LGCFMRS)
traceplot(LGCFMRS.mcmc)
autocorr.plot(LGCFMRS.mcmc)
```

And for the LGC model with a random intercept and a random slope the code becomes,

```
initsLGCRMRS <- function(){
```



```
  list(mean.alpha=rnorm(1, mean=1000, sd=100), mean.beta=rnorm(1,
       mean=1000, sd=100))

  list(mean.alpha=rnorm(1, mean=1000, sd=100), mean.beta=rnorm(1,
       mean=1000,sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

LGCRMRS <- jags(data = list("DV","T","N","Occasion"),
           initsLGCRMRS, model.file="[Path To File]",
           parameters.to.save=c("mean.alpha","var.y",
           "mean.beta"), n.chains =2, n.iter = Itt, n.burnin=Bi,
            n.thin=1, DIC=T)


LGCRMRS <- autojags(LGCRMRS, n.iter = Itt, n.thin=1,
   n.update= 10, DIC=T)

LGCRMRS.mcmc <- as.mcmc(LGCRMRS)
traceplot(LGCRMRS.mcmc)
autocorr.plot(LGCRMRS.mcmc)
```

When both the intercept and the slope are allowed to vary across individuals, as is the case in this last model, this raises the possibility that these two model parameters are correlated with each other. An estimate for this correlation can be acquired by adding "cor" to the *parameters.to.save* argument.

4.2.3 QGC Model

The only difference between a LGC model and a QGC model is that the QGC model contains one additional parameter (*eta*) that describes the quadratic

trend in the data. Therefore, the main difference between the code for this model and the codes of the LGC model discussed above, is that the lists of starting values now have to contain this additional (quadratic) parameter, and that we will also ask for estimates of this parameter in the *parameters.to.save* argument. In addition, this additional parameter means that the number of variations of the QGC model is larger than for the LGC. For the LGC model, we provide code for 1) a LGC model with a fixed intercept and a fixed slope, 2) a LGC model with a random intercept and a fixed slope, 3) a LGC model with a fixed intercept and a random slope, and 4) a LGC model with a random intercept and a random slope. For the QGC model we provide code for 1) a QGC model with a fixed intercept, slope, and quadratic term 2) a QGC model with a random intercept and a fixed slope and quadratic term, 3) a QGC model with a fixed intercept and quadratic term, and a random slope, 4) a QGC model with a fixed intercept and slope, and a random quadratic term, 5) a QGC with a random intercept and slope, and a fixed quadratic term, 6) a QGC with a random intercept and quadratic term, and a fixed slope, 7) a QGC with a fixed intercept and a random slope and quadratic term, and 8) a QGC model in which all three parameters are random.

The code for the QGC model with all three parameters fixed is,

```
initsQGCFMFSFSQ <- function(){
  list(alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1,
       mean=1000, sd=100), eta=rnorm(1, mean=1000, sd=100))

  list(alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1,
       mean=1000, sd=100), eta=rnorm(1, mean=1000, sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

QGCFMFSFSQ <- jags(data = list("DV","T","N","Occasion"),
  initsQGCFMFSFSQ, model.file="[Path To File]",
```



```
  parameters.to.save= c("alpha","var.y", "beta",
  "eta"), n.chains =2, n.iter = Itt, n.burnin=Bi,
  n.thin=1, DIC=T)

QGCFMFSFSQ <- autojags(QGCFMFSFSQ, n.iter = Itt, n.thin=1,
  n.update= 10, DIC=T)

QGCFMFSFSQ.mcmc <- as.mcmc(QGCFMFSFSQ)
traceplot(QGCFMFSFSQ.mcmc)
autocorr.plot(QGCFMFSFSQ.mcmc)
```

The code for the QGC model with a random intercept is,

```
initsQGCRMFSFSQ <- function(){
  list(mean.alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100), eta=rnorm(1, mean=1000, sd=100))

  list(mean.alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100), eta=rnorm(1, mean=1000, sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

QGCRMFSFSQ <- jags(data = list("DV","T","N","Occasion"),
  initsQGCRMFSFSQ, model.file="[Path To File]",
  parameters.to.save= c("mean.alpha","var.y", "beta",
  "eta"), n.chains =2, n.iter = Itt, n.burnin=Bi,
  n.thin=1, DIC=T)

QGCRMFSFSQ <- autojags(QGCRMFSFSQ, n.iter = Itt, n.thin=1,
  n.update= 10, DIC=T)

QGCRMFSFSQ.mcmc <- as.mcmc(QGCRMFSFSQ)
```

```
traceplot(QGCRMFSFSQ.mcmc)
autocorr.plot(QGCRMFSFSQ.mcmc)
```

The code for the QGC model with a random slope is,

```
initsQGCFMRSFSQ <- function(){
  list(alpha=rnorm(1, mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100), eta=rnorm(1, mean=1000, sd=100))

  list(alpha=rnorm(1, mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100), eta=rnorm(1, mean=1000, sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

QGCFMRSFSQ <- jags(data = list("DV","T","N","Occasion"),
  initsQGCFMRSFSQ, model.file="[Path To File]",
  parameters.to.save= c("alpha","var.y", "mean.beta",
  "eta"), n.chains =2, n.iter = Itt, n.burnin=Bi,
  n.thin=1, DIC=T)

QGCFMRSFSQ <- autojags(QGCFMRSFSQ, n.iter = Itt, n.thin=1,
n.update= 10, DIC=T)

QGCFMRSFSQ.mcmc <- as.mcmc(QGCFMRSFSQ)
traceplot(QGCFMRSFSQ.mcmc)
autocorr.plot(QGCFMRSFSQ.mcmc)
```

The code for the QGC model with a random quadratic term is,

```
initsQGCFMFSRSQ <- function(){
  list(alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1, mean=1000,
  sd=100), mean.eta=rnorm(1, mean=1000, sd=100))
```



```
  list(alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1, mean=1000,
  sd=100), mean.eta=rnorm(1, mean=1000, sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

QGCFMFSRSQ <- jags(data = list("DV","T","N","Occasion"),
  initsQGCFMFSRSQ, model.file="[Path To File]",
  parameters.to.save= c("alpha","var.y", "beta",
  "mean.eta"), n.chains =2, n.iter = Itt, n.burnin=Bi,
  n.thin=1, DIC=T)

QGCFMFSRSQ <- autojags(QGCFMFSRSQ, n.iter = Itt, n.thin=1,
  n.update= 10, DIC=T)

QGCFMFSRSQ.mcmc <- as.mcmc(QGCFMFSRSQ)
traceplot(QGCFMFSRSQ.mcmc)
autocorr.plot(QGCFMFSRSQ.mcmc)
```

The code for the QGC model with a random intercept and slope is,

```
initsQGCRMRSFSQ <- function(){
  list(mean.alpha=rnorm(1, mean=1000, sd=100),
  mean.beta=rnorm(1, mean=1000, sd=100), eta=rnorm(1,
  mean=1000, sd=100))

  list(mean.alpha=rnorm(1, mean=1000, sd=100),
  mean.beta=rnorm(1, mean=1000, sd=100), eta=rnorm(1,
  mean=1000, sd=100))

}
```

```
DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

QGCRMRSFSQ <- jags(data = list("DV","T","N","Occasion"),
  initsQGCRMRSFSQ, model.file="[Path To File]",
  parameters.to.save= c("mean.alpha","var.y",
  "mean.beta", "eta"), n.chains =2, n.iter = Itt,
  n.burnin=Bi, n.thin=1, DIC=T)

QGCRMRSFSQ <- autojags(QGCRMRSFSQ, n.iter = Itt, n.thin=1,
  n.update= 10, DIC=T)

QGCRMRSFSQ.mcmc <- as.mcmc(QGCRMRSFSQ)
traceplot(QGCRMRSFSQ.mcmc)
autocorr.plot(QGCRMRSFSQ.mcmc)
```

The code for the QGC model with a random intercept and quadratic term is,

```
initsQGCRMFSRSQ <- function(){
  list(mean.alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100), mean.eta=rnorm(1, mean=1000, sd=100))

  list(mean.alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100), mean.eta=rnorm(1, mean=1000, sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

QGCRMFSRSQ <- jags(data = list("DV","T","N","Occasion"),
```



```
  initsQGCRMFSRSQ, model.file="[Path To File]",
  parameters.to.save= c("mean.alpha","var.y", "beta",
  "mean.eta"), n.chains =2, n.iter = Itt, n.burnin=Bi,
  n.thin=1, DIC=T)

QGCRMFSRSQ <- autojags(QGCRMFSRSQ, n.iter = Itt, n.thin=1,
n.update= 10, DIC=T)

QGCRMFSRSQ.mcmc <- as.mcmc(QGCRMFSRSQ)
traceplot(QGCRMFSRSQ.mcmc)
autocorr.plot(QGCRMFSRSQ.mcmc)
```

The code for the QGC model with a random slope and quadratic term is,

```
initsQGCFMRSRSQ <- function(){
  list(alpha=rnorm(1, mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100), mean.eta=rnorm(1, mean=1000, sd=100))

  list(alpha=rnorm(1, mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100), mean.eta=rnorm(1, mean=1000, sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

QGCFMRSRSQ <- jags(data = list("DV","T","N","Occasion"),
  initsQGCFMRSRSQ, model.file="[Path To File]",
  parameters.to.save= c("alpha","var.y", "mean.beta",
  "mean.eta"), n.chains =2, n.iter = Itt, n.burnin=Bi,
  n.thin=1, DIC=T)

QGCFMRSRSQ <- autojags(QGCFMRSRSQ, n.iter = Itt, n.thin=1,
  n.update= 10, DIC=T)
```

```
QGCFMRSRSQ.mcmc <- as.mcmc(QGCFMRSRSQ)
traceplot(QGCFMRSRSQ.mcmc)
autocorr.plot(QGCFMRSRSQ.mcmc)
```

Finally, the code for the QGC model in which all parameters are random is,

```
initsQGCRMRSRSQ <- function(){
  list(mean.alpha=rnorm(1, mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100), mean.eta=rnorm(1, mean=1000, sd=100))

  list(mean.alpha=rnorm(1, mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100), mean.eta=rnorm(1, mean=1000, sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

QGCRMRSRSQ <- jags(data = list("DV","T","N","Occasion"),
  initsQGCRMRSRSQ, model.file="[Path To File]",
  parameters.to.save= c("mean.alpha","var.y",
  "mean.beta", "mean.eta"), n.chains =2,
  n.iter = Itt, n.burnin=Bi, n.thin=1, DIC=T)

QGCRMRSRSQ <- autojags(QGCRMRSRSQ, n.iter = Itt, n.thin=1,
  n.update= 10, DIC=T)

QGCRMRSRSQ.mcmc <- as.mcmc(QGCRMRSRSQ)
traceplot(QGCRMRSRSQ.mcmc)
autocorr.plot(QGCRMRSRSQ.mcmc)
```

4.2.4 ALT Model

The final Bayesian multilevel model that can be fitted with our code is the ALT model. This model can be viewed as a combination of a LGC model and

a AR(1) model and the code for this model therefore also is a combination of the code for these two models. Specifically, the only difference between a LGC model and an ALT model is that the ALT model also contains an AR-parameter. Therefore, the biggest difference between the code for this model and the codes of the LGC model discussed above, is that the lists of starting values now also have to contain the AR parameter, and that we will also ask for estimates of this parameter in the *parameters.to.save* argument. Additionally however, the intercept and the slope of the growth curves of the ALT model are a function of the parameters *alfa*, *beta*, and the AR parameter. This means that the separate parameters *alfa* and *beta* do not have a clear interpretation in this model (Jongerling & Hamaker, 2011) and that we somehow have to model a combination of these parameters to get the intercept and the slope of the ALT model. We provided two ways of doing this by writing code for two different parametrizations of the ALT model. In the first parametrization we add code to the model file that expresses how the model parameters and the intercept and slope of the growth curve are related. In other words, in this parametrization we simply model the intercept and slope as a function of the *alfa*, *beta*, and AR parameter. In the second parametrization we rewrite the ALT model as a LGC model with autoregression between the measurement errors at successive time points. These 2 parametrizations are mathematically equivalent, but the second one separates the autoregressive part from the latent growth curve part (Hamaker, 2005) which means that in this parametrization of the ALT model, the separate model parameters *alfa*, *beta*, and *AR*, can simple be interpreted as the intercept, slope and AR-parameter, respectively.

Starting with the first parametrization of the ALT model, we provide code for 1) an ALT model with a fixed *alfa* and a fixed *beta* parameter, 2) an ALT model with a random *alfa* and a fixed *beta* parameter, 3) an ALT model with a fixed *alfa* and a random *beta* parameter, and 4) an ALT model with a random *alfa* and a random *beta* parameter. The AR parameter is always modeled as fixed because our data contained too few time points to reliably model intra-individual differences in this parameter.

The code for the ALT model with all three parameters fixed is,

```
initsALTFAFB <- function(){
  list(alpha=rnorm(1,mean=1000, sd=100), beta=rnorm(1,mean=1000,
  sd=100), AR=runif(1,0,1))
```

```
  list(alpha=rnorm(1,mean=1000, sd=100), beta=rnorm(1,mean=1000,
  sd=100), AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

ALTFAFB <- jags(data = list("DV","T","N","Occasion"),
   initsALTFAFB, model.file="[Path To File]",
   parameters.to.save= c("intercept","var.y", "slope",
   "AR"), n.chains =2, n.iter = Itt, n.burnin=Bi,
   n.thin=1, DIC=T)

ALTFAFB <- autojags(ALTFAFB, n.iter = Itt, n.thin=1,
   n.update= 10, DIC=T)


ALTFAFB.mcmc <- as.mcmc(ALTFAFB)
autocorr.plot(ALTFAFB.mcmc)
traceplot(ALTFAFB.mcmc)
```

As mentioned above the biggest difference with the code for the LGC model discussed above is that the lists of starting values (specified on the first 7 lines of code) now also contain the AR parameter, and that we ask for estimates of this parameter in the *parameters.to.save* argument. In addition, note that although the intercept and the slope are a function of several model parameters, it is the estimates of these 'composite' parameters that we request in the *parameters.to.save* argument.

The code for the ALT model with a random *alpha* parameter is,

```
initsALTRAFB <- function(){
  list(mean.alpha=rnorm(1,mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100), AR=runif(1,0,1))

  list(mean.alpha=rnorm(1,mean=1000, sd=100), beta=rnorm(1,
```



```
  mean=1000, sd=100), AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

ALTRAFB <- jags(data = list("DV","T","N","Occasion"),
   initsALTRAFB, model.file="[Path To File]",
   parameters.to.save= c("mean.intercept","var.y",
   "slope","AR"), n.chains =2, n.iter = Itt, n.burnin=Bi,
   n.thin=1, DIC=T)

ALTRAFB <- autojags(ALTRAFB, n.iter = Itt, n.thin=1, n.update= 10,
   DIC=T)


ALTRAFB.mcmc <- as.mcmc(ALTRAFB)
autocorr.plot(ALTRAFB.mcmc)
traceplot(ALTRAFB.mcmc)
```

Note that allowing *alpha* to vary across individuals results in the intercept of the growth curve to also be random. We therefore ask for the average of this composite parameter across individuals (*mean.intercept*) in the *parameters.to.save* argument.

The code for a ALT model with a random *beta* parameter is,

```
initsALTFARB <- function(){
  list(alpha=rnorm(1,mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100), AR=runif(1,0,1))

  list(alpha=rnorm(1,mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100), AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
```

```
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

ALTFARB <- jags(data = list("DV","T","N","Occasion"),
   initsALTFARB, model.file="[Path To File]",
   parameters.to.save= c("mean.intercept","var.y",
   "mean.slope","AR"), n.chains =2, n.iter = Itt,
   n.burnin=Bi, n.thin=1, DIC=T)

ALTFARB <- autojags(ALTFARB, n.iter = Itt, n.thin=1, n.update= 10,
   DIC=T)


ALTFARB.mcmc <- as.mcmc(ALTFARB)
autocorr.plot(ALTFARB.mcmc)
traceplot(ALTFARB.mcmc)
```

Note that allowing *beta* to vary across individuals results in both the intercept and the slope of the growth curve to also be random. We therefore ask for the averages of both these composite parameters (*mean.intercept* and *mean.slope*) in the *parameters.to.save* argument.

Finally, the code for the ALT model with both a random *alpha* and a random *beta* parameter is,

```
initsALTRARB <- function(){
  list(mean.alpha=rnorm(1,mean=1000, sd=100),
  mean.beta=rnorm(1,mean=1000, sd=100),
  AR=runif(1,0,1))

  list(mean.alpha=rnorm(1,mean=1000, sd=100),
  mean.beta=rnorm(1,mean=1000, sd=100),
  AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
```



```
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

ALTRARB <- jags(data = list("DV","T","N","Occasion"),
   initsALTRARB, model.file="[Path To File]",
   parameters.to.save= c("mean.intercept","var.y",
   "mean.slope","AR"), n.chains =2, n.iter = Itt,
   n.burnin=Bi, n.thin=1, DIC=T)

ALTRARB <- autojags(ALTRARB, n.iter = Itt, n.thin=1, n.update= 10,
   DIC=T)


ALTRARB.mcmc <- as.mcmc(ALTRARB)
autocorr.plot(ALTRARB.mcmc)
traceplot(ALTRARB.mcmc)
```

Like an ALT model with a random *beta* parameter, an ALT model with both a random *alpha* and a random *beta* parameter has both a random intercept and a random slope. We therefore ask for the averages of both these composite parameters (*mean.intercept* and *mean.slope*) in the *parameters.to.save* argument. In this variation of the ALT model however, the inta-individual variance in the slope is less constrained than in an ALT model with just a random *beta* parameter.

For the second parametrization of the ALT model, in which we rewrite the ALT model as a LGC model with autoregression between the measurement errors at successive time points, we provide code for 1) an ALT model with a fixed intercept and a fixed slope, 2)an ALT model with a random intercept and a fixed slope, 3) an ALT model with a fixed intercept and a random slope, and 4) an ALT model with a random intercept and a random slope. We again keep the AR parameter fixed in all variations of the ALT model because our data contained too few time points to reliably model intra-individual differences in this parameter.

The code for these 4 variations of the ALT model is almost identical to that of the 4 variations of the model under the first paramerization. However, in this parametrization the interpretation of the model parameters was straightforward, and the *alpha* and *beta* parameter can once again be inter-

preted as the intercept and slope of the growth curve. The only difference with the code for the 4 variations of the ALT model under the first specification is therefore that we directly ask for estimates of (the average values of) these two model parameters under the *parameters.to.save* argument, and not for the composite parameters we used above.

The code for the ALT model with a fixed intercept, a fixed slope, and fixed AR parameter then becomes,

```
initsALTFMFS <- function(){
  list(alpha=rnorm(1,mean=1000, sd=100), beta=rnorm(1,mean=1000,
  sd=100), AR=runif(1,0,1))

  list(alpha=rnorm(1,mean=1000, sd=100), beta=rnorm(1,mean=1000,
  sd=100), AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

ALTFMFS <- jags(data = list("DV","T","N","Occasion"),
   initsALTFMFS, model.file="[Path To File]",
   parameters.to.save= c("alpha","var.y", "beta", "AR"),
   n.chains =2, n.iter = Itt, n.burnin=Bi, n.thin=1,
   DIC=T)

ALTFMFS <- autojags(ALTFMFS, n.iter = Itt, n.thin=1, n.update= 10,
   DIC=T)


ALTFMFS.mcmc <- as.mcmc(ALTFMFS)
autocorr.plot(ALTFMFS.mcmc)
traceplot(ALTFMFS.mcmc)
```

The code for the ALT model with a random intercept becomes is,

```
initsALTRMFS <- function(){
```



```
  list(mean.alpha=rnorm(1,mean=1000, sd=100),
  beta=rnorm(1,mean=1000, sd=100), AR=runif(1,0,1))

  list(mean.alpha=rnorm(1,mean=1000, sd=100),
  beta=rnorm(1,mean=1000, sd=100), AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

ALTRMFS <- jags(data = list("DV","T","N","Occasion"),
   initsALTRMFS, model.file="[Path To File]",
   parameters.to.save= c("mean.alpha","var.y", "beta",
   "AR"), n.chains =2, n.iter = Itt, n.burnin=Bi, n.thin=1,
   DIC=T)

ALTRMFS <- autojags(ALTRMFS, n.iter = Itt, n.thin=1, n.update= 10,
   DIC=T)


ALTRMFS.mcmc <- as.mcmc(ALTRMFS)
autocorr.plot(ALTRMFS.mcmc)
traceplot(ALTRMFS.mcmc)
```

The code for a ALT model with a random slope is,

```
initsALTFMRS <- function(){
  list(alpha=rnorm(1,mean=1000, sd=100),
  mean.beta=rnorm(1,mean=1000, sd=100), AR=runif(1,0,1))

  list(alpha=rnorm(1,mean=1000, sd=100),
  mean.beta=rnorm(1,mean=1000, sd=100), AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
```

```
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

ALTFMRS <- jags(data = list("DV","T","N","Occasion"),
   initsALTFMRS, model.file="[Path To File]",
   parameters.to.save= c("alpha","var.y", "mean.beta",
   "AR"), n.chains =2, n.iter = Itt, n.burnin=Bi,
   n.thin=1, DIC=T)

ALTFMRS <- autojags(ALTFMRS, n.iter = Itt, n.thin=1, n.update= 10,
   DIC=T)


ALTFMRS.mcmc <- as.mcmc(ALTFMRS)
autocorr.plot(ALTFMRS.mcmc)
traceplot(ALTFMRS.mcmc)
```

Finally, the code for the ALT model with both a random intercept and a random slope is,

```
initsALTRMRS <- function(){
  list(mean.alpha=rnorm(1,mean=1000, sd=100),
  mean.beta=rnorm(1,mean=1000, sd=100), AR=runif(1,0,1))

  list(mean.alpha=rnorm(1,mean=1000, sd=100),
  mean.beta=rnorm(1,mean=1000, sd=100), AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

ALTRMRS <- jags(data = list("DV","T","N","Occasion"),
   initsALTRMRS, model.file="[Path To File]",
```



```
   parameters.to.save= c("mean.alpha","var.y",
   "mean.beta","AR"), n.chains =2, n.iter = Itt, n.burnin=Bi,
   n.thin=1, DIC=T)

ALTRMRS <- autojags(ALTRMRS, n.iter = Itt, n.thin=1, n.update= 10,
   DIC=T)


ALTRMRS.mcmc <- as.mcmc(ALTRMRS)
autocorr.plot(ALTRMRS.mcmc)
traceplot(ALTRMRS.mcmc)
```

5 QGC Model with Group as Predictor for the Random Parameters

In our applied example the QGC model with a random intercept and slope had the best model fit. A logical next step therefore was adding predictors for the inter-individual differences in the intercept, the slope, or both. As a predictor we used group membership as indicated by the two dummy variables created when reading the data into R. Adding Group as a predictor for the random intercept implies that we can write the intercept α_i as,

$$\alpha_i = \gamma_{00} + \gamma_{10} * D1_i + \gamma_{20} * D2_i + \varepsilon_{\alpha i}, \tag{1}$$

where γ_{00} is the intercept for the reference group, γ_{10} is the regression coefficient for *D1* which quantifies the difference in intercept between the reference group and the group belonging to a score of 1 on *D1*, and γ_{20} is the regression coefficient for *D2* which quantifies the difference in intercept between the reference group and the group belonging to a score of 1 on *D2*. The term $\varepsilon_{\alpha i}$ is a random error term. Since the intercept is now a function of the three parameters γ_{00}, γ_{10}, and γ_{20} we ask for these parameters in the *parameters.to.save* argument instead of asking for α. This makes the code for the QGC model with Group as a predictor for the intercept equal to,

```
initsQGCRMRSFSQPredM <- function(){
  list(gamma00=rnorm(1, mean=1000, sd=100), gamma10=rnorm(1,
  mean=1000, sd=100), gamma20=rnorm(1, mean=1000, sd=100),
```

```
  mean.beta=rnorm(1, mean=1000, sd=100), eta=rnorm(1, mean=1000,
  sd=100))

  list(gamma00=rnorm(1, mean=1000, sd=100), gamma10=rnorm(1,
  mean=1000, sd=100), gamma20=rnorm(1, mean=1000, sd=100),
  mean.beta=rnorm(1, mean=1000, sd=100), eta=rnorm(1, mean=1000,
  sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]


QGCRMRSFSQPredM <- jags(data = list("DV","T","N","Occasion",
   "D1", "D2"), initsQGCRMRSFSQPredM,
   model.file="[Path To File]",
   parameters.to.save=c("gamma00","gamma10",
   "gamma20","var.y", "Cov", "mean.beta",
   "eta"), n.chains =2, n.iter = Itt,
   n.burnin=Bi, n.thin=1, DIC=T)


QGCRMRSFSQPredM <- autojags(QGCRMRSFSQPredM, n.iter = Itt, n.thin=1,
                   n.update= 10, DIC=T)


QGCRMRSFSQPredM.mcmc <- as.mcmc(QGCRMRSFSQPredM)
autocorr.plot(QGCRMRSFSQPredM.mcmc)
traceplot(QGCRMRSFSQPredM.mcmc)
```

When we add Group as a predictor for the random slope, the parameter β_i can be written as,

$$\beta_i = \gamma_{01} + \gamma_{11} * D1_i + \gamma_{21} * D2_i + \varepsilon_{\beta i}, \tag{2}$$

where γ_{01} is the slope for the reference group, γ_{11} is the regression coefficient for *D1* which quantifies the difference in slope between the reference group and the group belonging to a score of 1 on *D1*, and γ_{21} is the regression coefficient for *D2* which quantifies the difference in slope between the reference group and the group belonging to a score of 1 on *D2*. The term $\varepsilon_{\beta i}$ is a random error term. Since the slope is now a function of the three parameters γ_{01}, γ_{11}, and γ_{21} we again ask for these parameters in the *parameters.to.save* argument instead of asking for β. This makes the code for the QGC model with Group as a predictor for the slope equal to,

```
initsQGCRMRSFSQPredS <- function(){
  list(gamma01=rnorm(1, mean=1000, sd=100), gamma11=rnorm(1,
  mean=1000, sd=100), gamma21=rnorm(1, mean=1000, sd=100),
  mean.alpha=rnorm(1, mean=1000, sd=100), eta=rnorm(1,
  mean=1000, sd=100))


  list(gamma01=rnorm(1, mean=1000, sd=100), gamma11=rnorm(1,
  mean=1000, sd=100), gamma21=rnorm(1, mean=1000, sd=100),
  mean.alpha=rnorm(1, mean=1000, sd=100), eta=rnorm(1,
  mean=1000, sd=100))

}

QGCRMRSFSQPredS <- jags(data = list("DV","T","N","Occasion",
   "D1", "D2"), initsQGCRMRSFSQPredS,
   model.file="[Path To File]",
                   parameters.to.save=c("gamma01","gamma11",
                   "gamma21", "var.y", "Cov", "mean.alpha",
                   "eta"),n.chains =2, n.iter = Itt,
                   n.burnin=Bi, n.thin=1, DIC=T)


QGCRMRSFSQPredS <- autojags(QGCRMRSFSQPredS, n.iter = Itt, n.thin=1,
                   n.update= 10, DIC=T)


QGCRMRSFSQPredS.mcmc <- as.mcmc(QGCRMRSFSQPredS)
```

```
autocorr.plot(QGCRMRSFSQPredS.mcmc)
traceplot(QGCRMRSFSQPredS.mcmc)
```

Finally, the code for a QGC model in which Group is a predictor for both the intercept and the slope can be written as,

```
initsQGCRMRSFSQPredMS <- function(){
  list(gamma01=rnorm(1, mean=1000, sd=100), gamma11=rnorm(1,
  mean=1000, sd=100), gamma21=rnorm(1, mean=1000, sd=100),
  gamma00=rnorm(1, mean=1000, sd=100), gamma10=rnorm(1,
  mean=1000, sd=100), gamma20=rnorm(1, mean=1000, sd=100),
  eta=rnorm(1, mean=1000, sd=100))

  list(gamma01=rnorm(1, mean=1000, sd=100), gamma11=rnorm(1,
  mean=1000, sd=100), gamma21=rnorm(1, mean=1000, sd=100),
  gamma00=rnorm(1, mean=1000, sd=100), gamma10=rnorm(1,
  mean=1000, sd=100), gamma20=rnorm(1, mean=1000, sd=100),
  eta=rnorm(1, mean=1000, sd=100))

}

QGCRMRSFSQPredMS <- jags(data = list("DV","T","N","Occasion",
"D1", "D2"), initsQGCRMRSFSQPredMS,
model.file="[Path To File]",
                    parameters.to.save=c("gamma00","gamma10",
                    "gamma20", "gamma01","gamma11", "gamma21",
                    "var.y", "Cov", "eta", "PredCon","PredST",
                    "PredCCT", "CEACST", "CEACCCT", "CEACCon"),

                    n.chains =2, n.iter = Itt, n.burnin=Bi,
                    n.thin=1, DIC=T)


QGCRMRSFSQPredMS <- autojags(QGCRMRSFSQPredMS, n.iter = Itt, n.thin=1,
                    n.update= 10, DIC=T)

QGCRMRSFSQPredMS.mcmc <- as.mcmc(QGCRMRSFSQPredMS)
autocorr.plot(QGCRMRSFSQPredMS.mcmc)
traceplot(QGCRMRSFSQPredMS.mcmc)
```


Adding second level predictors to the (random) parameters of models other than the QGC model can be done in a similar manner. One just has to rewrite the mean parameter value across all individuals as a linear relation of the predictor(s) that one wants to add, and subsequently include the regression coefficients for these second level predictors in the *parameters.to.save* argument. Further instructions on adding second level predictors are given in Appendix B of the article by Wetzelaer, Jongerling, Arntz, and Evers (2017).

Finally, applied researchers might be interested in comparing different groups in terms of their total scores on the dependent variable *DV* (i.e. total net benefit). This can easily be done with our Bayesian models by adding a set of variables to the JAGS model files that represent the sum scores of the *DV* across the different assessments, for each individual group (see the section "*QGC Random Intercept and Slope, Fixed Quadratic Term: Predictor for Intercept and Slope*" in Appendix B). In the code above for example, we added 6 additional statements to the *parameters.to.save* argument; *PredCon*, *PredST*, *PredCCT*, *CEACST*, *CEACCCT*, and *CEACCon*. The statements *PredCon*, *PredST*, and *PredCCT* are vectors containing the predicted scores on the *DV* for the *Con*, *ST*, and *CCT* group, at each of the assessments. The posterior distributions of these total scores for the different groups can now be used to draw conclusions about their relative performance.

The comparison of groups in terms of their total scores can be made even easier with the creation of (further) additional variables that directly represent the difference in total score between groups (e.g. incremental total net benefit). Note that for total NB we exclude the first (baseline) assessment since total NB should only be based on the costs after treatment has started. Variables that represent the incremental total net benefit between two groups are easily created by subtracting the total NB of one group from that of another. The variables *INBSTvsCon*, *INBSTvsCCT*, *INBCCTvsCon*, *INBConvsST*, *INBCCTvsST*, and *INBConvsCCT* in section "*QGC Random Intercept and Slope, Fixed Quadratic Term: Predictor for Intercept and Slope*" of Appendix B for instance, are examples of such variables that can be used to directly evaluate the incremental total net benefit between groups.

The probabilities for the Cost-Effectiveness Acceptability Curves (CEACs) can then be estimated by making use of the step function in JAGS for the variables *INBSTvsCon*, *INBSTvsCCT*, *INBCCTvsCon*, *INBConvsST*, *INBCCTvsST*, and *INBConvsCCT*. This step function takes the value 1 for each iteration where the value inside the parentheses is greater than zero. The variables *CEACST*, *CEACCCT*, and *CEACCon* in section "*QGC Ran-*

dom Intercept and Slope, Fixed Quadratic Term: Predictor for Intercept and Slope" of Appendix B therefore represent the probability that the total net benefit of one treatment is greater than that of the other two, or, in other words, the probability of that treatment being the most cost-effective of the three.

6 Analyzing Data with Three Levels

Since clinical trials will often take place in several mental health centres, and these centres can be seen as an additional level in our hierarchical data (i.e. observations are nested within individuals, and these individuals are nested in the different centres), we also provide code for the analysis of a QGC model with a random intercept and a random slope, that includes such a third (centre) level.

```
Centre <- [Column of the Data Matrix That Contains the Centre
           In Which The Individual Participated]
C <- [Number of Third Level Units]

initsQGCRMRSFSQL3 <- function(){
  list(mean.alpha=rnorm(C, mean=1000, sd=100),
  mean.beta=rnorm(C, mean=1000, sd=100),
  eta=rnorm(1, mean=1000, sd=100))

  list(mean.alpha=rnorm(C, mean=1000, sd=100),
  mean.beta=rnorm(C, mean=1000, sd=100),
  eta=rnorm(1, mean=1000, sd=100))

}

QGCRMRSFSQL3 <- jags(data = list("DV","T","N","Occasion",
"Centre","C"), initsQGCRMRSFSQL3,
model.file="[Path To File]",
parameters.to.save= c("mean.alpha","var.y",
                "Cov", "mean.beta", "eta", "MA","MB","VarA",
                "VarB"), n.chains =2, n.iter = Itt,
                n.burnin=Bi, n.thin=500, DIC=T)
```



```
QGCRMRSFSQL3 <- autojags(QGCRMRSFSQL3, n.iter = Itt, n.thin=500,
                n.update= 10, DIC=T)


QGCRMRSFSQL3.mcmc <- as.mcmc(QGCRMRSFSQL3)
autocorr.plot(QGCRMRSFSQL3.mcmc)
par(mfrow=c(3,3))
traceplot(QGCRMRSFSQL3.mcmc)
```

The only difference between the code for this three level QGC model and the two-level QGC model presented earlier is that we need to specify a variable that indicates to which third level unit an individual belongs *Centre*, and a variable that indicates the number of third level units *C*. In addition, the *mean.alpha* and *mean.beta* term in the *parameters.to.save* argument will now give a separate intercept and linear slope value for each of the third level units. To also get the overall mean and the variance of the intercept and (linear) slopes across the third level units, we now have to ask for MA, MB, VarA, and VarB in *parameters.to.save*.

6.1 Three Level QGC Model With Group as Predictor for the Random Parameters

Like the two level QGC model, the three level QGC model can be extended to include predictors for the individual intercepts and slopes. As before, we use group membership (as indicated by the two dummy variables created when reading the data into R) as a predictor for the inter-individual differences in the intercepts and slopes, with the difference that there now is an additional level above the individual level on which we are modeling these inter-individual differences. When we added Group as a predictor for the random intercept and slope in the two level QGC model we showed α_i and β_i were then written as,

$$\alpha_i = \gamma_{00} + \gamma_{10} * D1_i + \gamma_{20} * D2_i + \varepsilon_{\alpha i}, \quad \text{and} \tag{3}$$
$$\beta_i = \gamma_{01} + \gamma_{11} * D1_i + \gamma_{21} * D2_i + \varepsilon_{\beta i}, \tag{4}$$

where γ_{00} and γ_{01} are the intercept and slope for the reference group, γ_{10} and γ_{11} are the regression coefficients for *D1* which respectively quantify 1) the difference in the intercepts, and 2) the difference in the slopes between the reference group and the group belonging to a score of 1 on *D1*. Finally, γ_{20} and γ_{21} are the regression coefficients for *D2* which respectively quantify 1) the difference in the intercepts, and 2) the difference in the slopes between the reference group and the group belonging to a score of 1 on *D2*. The terms $\varepsilon_{\alpha i}$ and $\varepsilon_{\beta i}$ are random error terms.

When we add a third level to the model the parameters γ_{00}, γ_{10}, γ_{20}, γ_{01}, γ_{11}, and γ_{21} can vary across the third level units, and we need to take this random variance on the third level into account. This is simply done by drawing a separate value for each of these six parameters, for each of the third level units, from a overarching distribution. Assuming normal distributions this implies,

$$\gamma_{00} \sim \mathcal{N}(\mu_{\gamma_{00}}, \tau^2_{\gamma_{00}}) \tag{5}$$
$$\gamma_{10} \sim \mathcal{N}(\mu_{\gamma_{10}}, \tau^2_{\gamma_{10}}) \tag{6}$$
$$\gamma_{20} \sim \mathcal{N}(\mu_{\gamma_{20}}, \tau^2_{\gamma_{20}}) \tag{7}$$
$$\gamma_{01} \sim \mathcal{N}(\mu_{\gamma_{01}}, \tau^2_{\gamma_{01}}) \tag{8}$$
$$\gamma_{11} \sim \mathcal{N}(\mu_{\gamma_{11}}, \tau^2_{\gamma_{11}}) \tag{9}$$
$$\gamma_{21} \sim \mathcal{N}(\mu_{\gamma_{21}}, \tau^2_{\gamma_{21}}), \tag{10}$$

where γ_{00}, γ_{10}, γ_{20}, γ_{01}, γ_{11}, and γ_{21} are vectors containing the separate parameter values for each higher level unit c; $\mu_{\gamma_{00}}$, $\mu_{\gamma_{10}}$, $\mu_{\gamma_{20}}$, $\mu_{\gamma_{01}}$, $\mu_{\gamma_{11}}$, and $\mu_{\gamma_{21}}$ are the overall means of these parameters, and $\tau^2_{\gamma_{00}}$, $\tau^2_{\gamma_{10}}$, $\tau^2_{\gamma_{20}}$, $\tau^2_{\gamma_{01}}$, $\tau^2_{\gamma_{11}}$, and $\tau^2_{\gamma_{21}}$ are the random variances in these parameters across the third level units. Subsequently, α_i and β_i can be written as,

$$\alpha_i = \gamma_{00}[c] + \gamma_{10}[c] * D1_i + \gamma_{20}[c] * D2_i + \varepsilon_{\alpha i}, \quad \text{and} \tag{11}$$
$$\beta_i = \gamma_{01}[c] + \gamma_{11}[c] * D1_i + \gamma_{21}[c] * D2_i + \varepsilon_{\beta i}, \tag{12}$$

where c between square brackets implies that we are selecting that value from the parameter vector that belongs to the higher level unit (c) to which an individual belongs.

The code to run this three-level model is very similar to that of the two-level QGC model with Group as a predictor for the random variances. It can be written as,

```
initsQGCRMRSFSQL3P <- function(){
  list(Mugamma00=rnorm(1, mean=1000, sd=100),
  Mugamma10=rnorm(1, mean=1000, sd=100),
  Mugamma20=rnorm(1, mean=1000, sd=100),
  Mugamma01=rnorm(1, mean=1000, sd=100),
  Mugamma11=rnorm(1, mean=1000, sd=100),
  Mugamma21=rnorm(1, mean=1000, sd=100),
  eta=rnorm(1, mean=1000, sd=100))

  list(Mugamma00=rnorm(1, mean=1000, sd=100),
  Mugamma10=rnorm(1, mean=1000, sd=100),
  Mugamma20=rnorm(1, mean=1000, sd=100),
  Mugamma01=rnorm(1, mean=1000, sd=100),
  Mugamma11=rnorm(1, mean=1000, sd=100),
  Mugamma21=rnorm(1, mean=1000, sd=100),
  eta=rnorm(1, mean=1000, sd=100))
  }

QGCRMRSFSQL3P <- jags(data = list("DV","T","N","Occasion",
 "D1", "D2", "Centre", "C"),
 initsQGCRMRSFSQL3P,
 model.file="[Path To File]",
                parameters.to.save=c("Mugamma00","Mugamma10",
                "Mugamma20","Mugamma01","Mugamma11","Mugamma21",
                "var.y", "Cov", "eta"), n.chains =2,
                n.iter = Itt, n.burnin=Bi, n.thin=500, DIC=T)


QGCRMRSFSQL3P <- autojags(QGCRMRSFSQL3P, n.iter = Itt, n.thin=500,
                n.update= 10, DIC=T)


QGCRMRSFSQL3P.mcmc <- as.mcmc(QGCRMRSFSQL3P)
autocorr.plot(QGCRMRSFSQL3P.mcmc)
```

```
par(mfrow=c(3,3))
traceplot(QGCRMRSFSQL3P.mcmc)
```

The only difference is that asking for γ_{00}, γ_{10}, γ_{20}, γ_{01}, γ_{11}, and γ_{21} in the *parameters.to.save* argument will now give use separate values for each of the higher level units. To get the average values for these parameters across the higher level units we have to ask for $\mu_{\gamma_{00}}$, $\mu_{\gamma_{10}}$, $\mu_{\gamma_{20}}$, $\mu_{\gamma_{01}}$, $\mu_{\gamma_{11}}$, and $\mu_{\gamma_{21}}$ instead.

7 Analyzing Lognormal or Gamma distributed Data

The R-code presented above assumes that the NB data is normally distributed, but our code can also be applied to lognormal and gamma distributed data. In order to apply a QGC model to lognormal data, one uses the logarithm of the data as the dependent variable. Note that this implies that also starting values and prior distributions need to be adjusted to the log scale. The following code can be used to fit a QGC model with a random intercept, random slope and fixed quadratic term to lognormal NB data,

```
DV <- ((-1*DV)+(lambda/6))
DV <- DV + 5
DV <- log(DV)


initsQGCRMRSFSQPredMSlog <- function(){
  list(gamma01=rnorm(1, mean=6.91, sd=4.61), gamma11=rnorm(1,
   mean=6.91, sd=4.61), gamma21=rnorm(1, mean=6.91, sd=4.61),
   gamma00=rnorm(1, mean=6.91, sd=4.61), gamma10=rnorm(1,
   mean=6.91, sd=4.61), gamma20=rnorm(1, mean=6.91, sd=4.61),
   eta=rnorm(1, mean=6.91, sd=4.61))

  list(gamma01=rnorm(1, mean=6.91, sd=4.61), gamma11=rnorm(1,
  mean=6.91, sd=4.61), gamma21=rnorm(1, mean=6.91, sd=4.61),
  gamma00=rnorm(1, mean=6.91, sd=4.61), gamma10=rnorm(1,
  mean=6.91, sd=4.61), gamma20=rnorm(1, mean=6.91, sd=4.61),
  eta=rnorm(1, mean=6.91, sd=4.61))
```



```
  }

QGCRMRSFSQPredMSlog <- jags(data = list("DV","T","N","Occasion",
   "D1", "D2"), initsQGCRMRSFSQPredMSlog,
   model.file="[Path To File]",
                       parameters.to.save=c("gamma00","gamma10",
                       "gamma20", "gamma01","gamma11", "gamma21",
                       "var.y", "Cov", "eta", "PredCon", "PredST", "PredCCT",
   "CEACST", "CEACCCT", "CEACCon"), n.chains =2,
                       n.iter = Itt, n.burnin=Bi, n.thin=1, DIC=T)

QGCRMRSFSQPredMSlog <- autojags(QGCRMRSFSQPredMSlog, n.iter = Itt,
   n.thin=1, n.update= 10, DIC=T)


QGCRMRSFSQPredMSlog.mcmc <- as.mcmc(QGCRMRSFSQPredMSlog)
autocorr.plot(QGCRMRSFSQPredMSlog.mcmc)
traceplot(QGCRMRSFSQPredMSlog.mcmc)
```

A lognormal distribution can only model data in which every score is larger than 0, but the NB data may contain negative values. In our dataset for example, the NB for a specific assessment of an individual who did not recover at follow-up will be minus the costs incurred during the time period that corresponds to that assessment. Therefore, in the first line of code the DV is multiplied with -1 to make the negative values positive. However, the NB values that were previously positive will now be negative. Since these are the values of patients who did recover at follow-up, those values can never exceed the value of lambda/6 that was used in calculating the NB. Hence, by also adding this value of lambda/6 to all values of NB, the first line of code effectively makes sure that none of the NB data is negative. The second line subsequently makes sure that there are no NB values equal to 0 by adding a small constant to the data. After making sure that all scores on the DV are larger than 0, we simply take the logarithm of the variable *DV*. The rest of the code is similar to that of the code used for normally distributed data. Only the starting values are adjusted to the new range of scores on the DV. Furthermore, researchers need to keep in mind that the resulting estimates will now also be on the log scale so that exponentiation is necessary for

interpretation of the results on the original scale.[3]

Like a lognormal distribution, a gamma distribution can only be fitted to data that does not contain values of 0 or lower. Therefore the first two lines of code are again to remove all values smaller than or equal to 0 from DV. Models for gamma distributed data are usually estimated by making use of a so-called log link-function, which we also included in our example below. This again implies that the starting values (and prior distribution for the covariance matrix) need to be adjusted to the log scale and that exponentiation of the resulting estimates will be necessary to enable interpretation on the original scale. Also note that since we multiplied the DV by -1, that also the calculation of the incremental total net benefit between treatments needs to be reversed (e.g. by calculating it the same way as before and then multiplying it by -1). The rest of the R-code used for gamma distributed data is the same as for normally distributed data, except that we now need to change the specified distribution of the data in the model file used by the JAGS program, such that,

$$y_{it} \sim \text{Gamma}(\kappa, \nu_{it}) \quad \text{and,} \tag{13}$$

$$\nu_{it} = \kappa/[\text{Mean Trend of Model}]. \tag{14}$$

where κ and ν are the shape and rate parameters of the Gamma distribution, and Equation 14 makes sure that the mean trend of the gamma distributed data follows the mean trend of the fitted model (since the mean of a gamma distribution is equal to $\frac{\mu_\kappa}{\mu_\nu}$). For a QGC model in which all three model parameters are random Equation 14 is equal to,

$$\nu_{it} = \kappa/\left(\delta_i + \gamma_i(t-1) + \theta_i(t-1)^2\right), \tag{15}$$

for example.

The code to fit a QGC model with a random intercept, random slope and fixed quadratic term to gamma distributed NB data is,

```
DV <- ((-1*DV)+(lambda/6))
DV <- DV + .0000001
```

[3]Please note that an additional Smearing correction is required for the correct interpretation of the results with lognormal data, see e.g. the model code in section 10.30.


```
initsQGCRMRSFSQPredMSgamma <- function(){

list(gamma01=rnorm(1, 6.91, 4.61), gamma11=rnorm(1, 6.91, 4.61),
gamma21=rnorm(1, 6.91, 4.61), gamma00=rnorm(1, 6.91, 4.61),
gamma10=rnorm(1, 6.91, 4.61), gamma20=rnorm(1, 6.91, 4.61),
eta=rnorm(1, 6.91, 4.61), shape=runif(1, 0, 20))

list(gamma01=rnorm(1, 6.91, 4.61), gamma11=rnorm(1, 6.91, 4.61),
gamma21=rnorm(1, 6.91, 4.61), gamma00=rnorm(1, 6.91, 4.61),
gamma10=rnorm(1, 6.91, 4.61), gamma20=rnorm(1, 6.91, 4.61),
eta=rnorm(1, 6.91, 4.61), shape=runif(1, 0, 20))

}


QGCRMRSFSQPredMSgamma <- jags(data = list("DV","T","N","Occasion",
                      "D1", "D2"),
                      initsQGCRMRSFSQPredMSgamma,
                      model.file="[Path To File]",
                      parameters.to.save=c("gamma00","gamma10",
                      "gamma20", "gamma01","gamma11", "gamma21",
                      "Cov", "eta", "shape", "PredCon", "PredST", "PredCCT"
   "CEACST", "CEACCCT", "CEACCon"),
   n.chains =2, n.iter = Itt, n.burnin=Bi, n.thin=1, DIC=T)

QGCRMRSFSQPredMSgamma <- autojags(QGCRMRSFSQPredMSgamma,
 n.iter = Itt, n.thin=1, n.update= 10,
 DIC=T)


QGCRMRSFSQPredMSgamma.mcmc <- as.mcmc(QGCRMRSFSQPredMSgamma)
autocorr.plot(QGCRMRSFSQPredMSgamma.mcmc)
traceplot(QGCRMRSFSQPredMSgamma.mcmc)
```

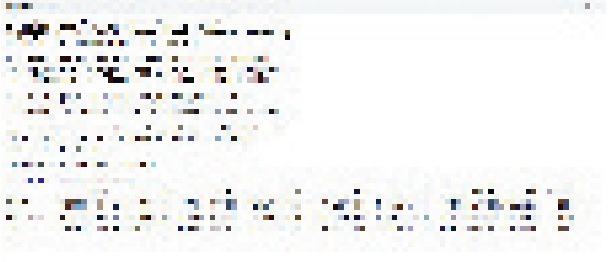
In Appendix B, a JAGS model file for the analysis of gamma distributed data with a QGC model with a random intercept, random slope and fixed

quadratic term is given. For the other models, the model files can be adapted for gamma distributed data by changing the specified distribution of DV, and setting the rate parameter $\nu_{i,t}$ equal to the shape parameter κ divided by the appropriate mean trend. For an example of how models for gamma distributed data can be extended to accommodate hierarchical data with additional levels, the interested reader is referred to (Thompson, Nixon, & Grieve, 2006).

8 Output

The code for each model provides the researcher with estimates for all the parameters specified in the *parameters.to.save* argument of the call to the JAGS program (Spiegelhalter et al., 2002). As an example, the output for the LGC-model with a fixed intercept and slope is given in Figure 3 As this figure shows, the JAGS program will return 1) the mean value, 2) the standard deviation, 3) the 2.5th, 25th, 50th, 75th, and 97.5th percentiles, 4) the Gelman-Rubin statistic (*Rhat*) (Gelman et al., 2004, pg. 296-298) for each of the parameters specified, and 5) the effective number of independent draws (Gelman, Carlin, Stern, & Rubin, 2004, pg. 298-299). Most of the values are self explanatory, but applied researchers are probably less familiar with the Gelman-Rubin statistic.

Figure 3: Example of Output from the JAGS Program.



This statistic is a measure of model convergence, where convergence has to do with the fact that Bayesian inference is based on the posterior distribution of model parameter. Specifically, a Gibbs-sampler tries to construct an approximation of the posterior distribution of a parameter by repeatedly drawing values from the posterior distribution of a parameter, conditional on the value of all the other parameters in the model. All inference is subsequently based on this approximate posterior distribution. The mean or median of the distribution is used as a point estimate for example, while the standard deviation can be used as an equivalent of the standard error. It is therefore important that the Gibbs-sampler is indeed sampling form the correct distribution. Whether this is the case is indicated by the convergence of the Gibbs-sampler. If the sampler is indeed sampling from the (conditional) posterior distribution of a parameter we say the Gibbs-sampler converged, if it is not sampling from the correct posterior, we say the sampler did not converge. Although convergence can never be absolutely proven, there are a few checks that researchers can do to make sure that convergence was (at least) likely reached.

The first of these checks involves the Gelman-Rubin statistic (Gelman et al., 2004, pg. 296-298). If the value of this statistic is smaller than 1.1 for all estimated parameters, then that is an indication that the Gibbs-sampler probably converged. If it is larger than 1.1 convergence has not been reached, and a researcher should first try letting the Gibbs-sampler run for some additional iterations. This can be done by using the lines of code starting with the *autojags* command, which will make the Gibbs-sampler run for additional iterations until the statistics for every parameter is smaller than 1.1. So for the AR(1) model with fixed mean, the following line of code

```
FixedAR <- autojags(FixedAR, n.iter = Itt, n.thin=1,
   n.update= 10, DIC=T),
```

can be run to ask for additional iterations. Most of the arguments of this call we already saw in the original call to the JAGS program, with two exceptions. The first exception is the first argument, which points to the variable in which a researcher stored his or her original MCMC-simulation. The autojags command tells JAGS to run *n.iter* new MCMC-simulations, starting with the parameter values of the last original simulation, and to check if all Gelman-Rubin statistics are smaller than 1.1 after they are completed. If so, the MCMC-simulation stops and the new set of simulations is stored in the

variable in the first argument of the autojags command. If the *n.iter* new MCMC-iterations did not result in Gelman-Rubin statistics smaller than 1.1, the autojags commands runs another set of *n.iter* new MCMC-simulations, starting with the parameter values calculated in the last simulation (i.e. the last simulation of the first set of additional simulations), and keeps doing so until *n.update* new sets of simulations have been run. This *n.update* argument is the second new argument in the autojags command.

Researchers should always check the Gelman-Rubin statistics for all estimated parameter to make sure that it is lower than 1.1, but as we already mentioned, this is only the first check of (probable) convergence. The second is plotting the repeated draws of the Gibbs-sampler using a so called traceplot. In a traceplot, the value that the Gibbs-sampler drew for a specific parameter is on the y-axis, while the iteration of the Gibbs-sampler is on the x-axis. If the Gibbs-sampler converged, then this traceplot should not show any trend, and the repeated draws of a parameter value should randomly fluctuate around a constant mean value. An example of a traceplot for a 2-chained converged Gibbs-sampler is given in Figure 4.

For all the models discussed above, these plots are made with the lines of code that turn the variable containing the JAGS output into an MCMC object, and subsequently plot this object using the *traceplot* command. So to make these plots for the AR(1) model with fixed mean, we would run the lines,

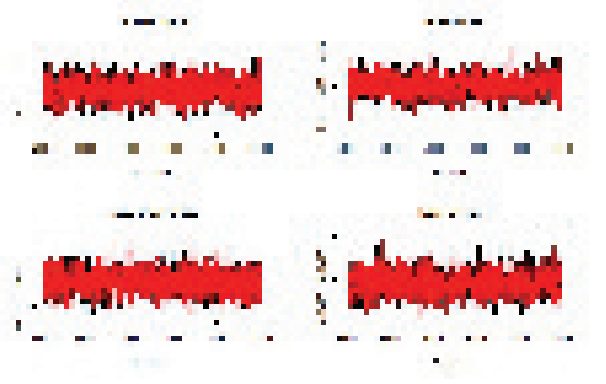
```
FixedAR.mcmc <- as.mcmc(FixedAR)
traceplot(FixedAR.mcmc)
```

Only when the Gelman-Rubin statistic for every estimated parameter is below 1.1 and the traceplots for these parameter show the random fluctuations illustrated in Figure 4 can we be fairly certain that convergence was reached.

In cases where the Gibbs-sampler does not converge, indicated by either Gelman-Rubin statistics above 1.1, traceplots that show something other than random fluctuations around a steady mean, or both, this could be due to very high correlation between successive draws from the Gibbs-sampler. To check if this is the case applied researchers can use the *auto.corr* command on the MCMC object created for the traceplots. For the fixed mean AR(1) model this would involve running the line,

```
autocorr.plot(FixedAR.mcmc).
```


Figure 4: Example of Traceplots for a Converged Model.



The outcome of this line of code is presented in Figure 5. The autocorrelation plot shows the amount of autocorrelations between draws on the y-axis, while the number of iterations between those draws is shown on the x-axis. Ideally, the plot should show autocorrelation that drops to 0 quickly (as is the case in Figure 5). If this is not the case, then it is recommended that researchers first increase the value of the *n.thin* argument so that intermediate draws from the Gibbs-sampler are discarded and the correlations between the draws that are kept are smaller than before. If increasing the thinning rate does not help, one should either use a reparametrization of the model (Gelman, Carlin, Stern, & Rubin, 2004, pg. 302-305), or use a different model altogether.

Finally, our approach will also provide the following output for the DIC (Spiegelhalter et al., 2002),

```
DIC info (using the rule, pD = var(deviance)/2)
pD = [pD Value for the Selected Model] and DIC = [DIC Value for the
  Selected Model]
DIC is an estimate of expected predictive error (lower deviance is
better).
```

Similar to other Information Criteria, like the AIC (Akaike, 1973) and BIC (Schwarz, 1978), the DIC (Spiegelhalter et al., 2002) is a measure of model (mis)fit based on a trade-off between model fit and model complexity. Here, model complexity is expressed as the number of effective model parameters by pD (Gelman, Carlin, Stern, & Rubin, 2004, pg. 182). Lower values on the DIC imply a 'better' model fit, and, as a rule of thumb, differences in DIC values larger than 5 are usually considered relevant.

Figure 5: Example of Autocorrelation plots for a Converged Model.

9 Appendix A: R-Code

The following R-code can be run in its entirety. However, given the importance of carefully monitoring convergence of the different models, I recommend running the code for one model at a time.

```
################## Reading in and preparing Data ##############
Yrepeated <- read.csv("[Path To File]", sep = "", header = TRUE,
                      na.strings = "999999.00")


########## Optional Code in Case Data is in Long format #######
YWidth <- matrix(NA,nrow=(nrow(Yrepeated))/6,ncol=6)
D1 <- matrix(NA,nrow=(nrow(Yrepeated))/6,ncol=1)
D2 <- matrix(NA,nrow=(nrow(Yrepeated))/6,ncol=1)
Yrec <- matrix(NA,nrow=(nrow(Yrepeated))/6,ncol=6)
ind <- unique(Yrepeated[,1])

for (i in 1:(nrow(Yrepeated)/6)){
YWidth[i,] <- Yrepeated[Yrepeated[,1]==ind[i],4]
}

for (i in 1:(nrow(Yrepeated)/6)){
D1[i,] <- mean(Yrepeated[Yrepeated[,1]==ind[i],2])
}

for (i in 1:(nrow(Yrepeated)/6)){
D2[i,] <- mean(Yrepeated[Yrepeated[,1]==ind[i],3])
}

for (i in 1:(nrow(Yrepeated)/6)){
    Yrec[i,] <- Yrepeated[Yrepeated[,1]==ind[i],5]
}

Yrepeated <-  matrix(cbind(ind,D1,D2,YWidth,Yrec),nrow=
              length(ind))
```

```
#############################################################

library(R2jags)

################ AR(1)Fixed Mean and AR ######################

initsFixedMuAR <- function(){
  list(alpha=rnorm(1), AR=runif(1,0,1))
  list(alpha=rnorm(1), AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]

FixedMuAR <- jags(data = list("DV","T","N"), initsFixedMuAR,
             model.file="[Path To File]", parameters.to.save=
             c("mu","var.y", "AR"), n.chains =2, n.iter = Itt,
             n.burnin=Bi,n.thin=1, DIC=TRUE).

FixedMuAR <- autojags(FixedMuAR, n.iter = Itt, n.thin=1,
 n.update= 10, DIC=T)

FixedMuAR.mcmc <- as.mcmc(FixedMuAR)
traceplot(FixedMuAR.mcmc)
autocorr.plot(FixedMuAR.mcmc)

################ AR(1)Random Mean and Fixed AR ###############
initsRandomMuFixedAR <- function(){
  list(mean.mu=rnorm(1), AR=runif(1,0,1))
  list(mean.mu=rnorm(1), AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]

FixedMuRandomAR <- jags(data = list("DV","T","N"),
```



```
   initsRandomMuFixedAR,
                model.file="[Path To File]",
                parameters.to.save= c("mean.mu","var.y",
                "AR"), n.chains =2, n.iter = Itt,
                n.burnin=Bi, n.thin=1, DIC=T)


FixedMuRandomAR <- autojags(FixedMuRandomAR, n.iter = Itt,
       n.thin=1, n.update= 10,DIC=T)

FixedMuRandomAR.mcmc <- as.mcmc(FixedMuRandomAR)
traceplot(FixedMuRandomAR.mcmc)
autocorr.plot(FixedMuRandomAR.mcmc)

############## AR(1) Fixed Mean and Random AR #################
initsFixedMuRandomAR <- function(){
  list(alpha=rnorm(1), mean.AR=runif(1,0,1))
  list(alpha=rnorm(1), mean.AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]

FixedMuRandomAR <- jags(data = list("DV","T","N"),
   initsFixedMuRandomAR,
                 model.file="[Path To File]",
                 parameters.to.save= c("mu","var.y",
                 "AR"), n.chains =2, n.iter = Itt,
                 n.burnin=Bi,n.thin=1, DIC=TRUE).

FixedMuRandomAR <- autojags(FixedAR, n.iter = Itt, n.thin=1,
   n.update= 10, DIC=T)

FixedMuRandomAR.mcmc <- as.mcmc(FixedMuRandomAR)
traceplot(FixedMuRandomAR.mcmc)
autocorr.plot(FixedMuRandomAR.mcmc)
```

```
######################## Random AR ############################
initsRandomAR <- function(){
  list(mean.mu=rnorm(1), AR=runif(1,0,1))
  list(mean.mu=rnorm(1), AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]

RandomAR <- jags(data = list("DV","T","N"), initsRandomAR,
            model.file="[Path To File]", parameters.to.save=
            c("mean.mu","var.y", "AR"), n.chains =2, n.iter =
            Itt, n.burnin=Bi, n.thin=1, DIC=T)


RandomAR <- autojags(RandomAR, n.iter = Itt, n.thin=1,
n.update= 10,DIC=T)

RandomAR.mcmc <- as.mcmc(RandomAR)
traceplot(RandomAR.mcmc)
autocorr.plot(RandomAR.mcmc)

############## LGC Fixed Intercept and Slope #################
initsLGCFMS <- function(){
  list(alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1, mean=
  1000, sd=100))

  list(alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1, mean=
  1000,sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]
```



```
LGCFMS <- jags(data = list("DV","T","N","Occasion"),
  initsLGCFMS, model.file="[Path To File]",
  parameters.to.save= c("alpha","var.y", "beta"),
  n.chains =2, n.iter = Itt, n.burnin=Bi, n.thin=1,
  DIC=T)

LGCFMS <- autojags(LGCFMS, n.iter = Itt, n.thin=50,
  n.update= 10, DIC=T)

LGCFMS.mcmc <- as.mcmc(LGCFMS)
traceplot(LGCFMS.mcmc)
autocorr.plot(LGCFMS.mcmc)

########## LGC Random Intercept, Fixed and Slope #############
initsLGCRMFS <- function(){
  list(mean.alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100))

  list(mean.alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

LGCRMFS <- jags(data = list("DV","T","N","Occasion"),
   initsLGCRMFS, model.file="[Path To File]",
   parameters.to.save= c("mean.alpha","var.y", "beta"),
   n.chains =2, n.iter = Itt, n.burnin=Bi, n.thin=1,
   DIC=T)


LGCRMFS <- autojags(LGCRMFS, n.iter = Itt, n.thin=1,
```

```
   n.update= 10, DIC=T)

LGCRMFS.mcmc <- as.mcmc(LGCRMFS)
traceplot(LGCRMFS.mcmc)
autocorr.plot(LGCRMFS.mcmc)

########## LGC Fixed Intercept, Random and Slope #############
initsLGCFMRS <- function(){
  list(alpha=rnorm(1, mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100))

  list(alpha=rnorm(1, mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

LGCFMRS <- jags(data = list("DV","T","N","Occasion"),
   initsLGCFMRS, model.file="[Path To File]",
   parameters.to.save= c("alpha","var.y", "mean.beta"),
   n.chains =2, n.iter = Itt, n.burnin=Bi, n.thin=1,
   DIC=T)


LGCFMRS <- autojags(LGCFMRS, n.iter = Itt, n.thin=1,
   n.update= 10, DIC=T)

LGCFMRS.mcmc <- as.mcmc(LGCFMRS)
traceplot(LGCFMRS.mcmc)
autocorr.plot(LGCFMRS.mcmc)

############## LGC Random Intercept and Slope ################
initsLGCRMRS <- function(){
```



```
  list(mean.alpha=rnorm(1, mean=1000, sd=100), mean.beta=
  rnorm(1, mean=1000, sd=100))

  list(mean.alpha=rnorm(1, mean=1000, sd=100), mean.beta=
  rnorm(1, mean=1000,sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

LGCRMRS <- jags(data = list("DV","T","N","Occasion"),
           initsLGCRMRS, model.file="[Path To File]",
           parameters.to.save=c("mean.alpha","var.y",
           "mean.beta"), n.chains =2, n.iter = Itt,
           n.burnin=Bi, n.thin=1, DIC=T)


LGCRMRS <- autojags(LGCRMRS, n.iter = Itt, n.thin=1,
   n.update= 10, DIC=T)

LGCRMRS.mcmc <- as.mcmc(LGCRMRS)
traceplot(LGCRMRS.mcmc)
autocorr.plot(LGCRMRS.mcmc)

######## QGC Fixed Intercept, Slope and Quadratic Term ########
initsQGCFMFSFSQ <- function(){
  list(alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1,
       mean=1000, sd=100), eta=rnorm(1, mean=1000, sd=100))

  list(alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1,
       mean=1000, sd=100), eta=rnorm(1, mean=1000, sd=100))

}
```

```
DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
     Occasions]

QGCFMFSFSQ <- jags(data = list("DV","T","N","Occasion"),
  initsQGCFMFSFSQ, model.file="[Path To File]",
  parameters.to.save= c("alpha","var.y", "beta",
  "eta"), n.chains =2, n.iter = Itt, n.burnin=Bi,
  n.thin=1, DIC=T)

QGCFMFSFSQ <- autojags(QGCFMFSFSQ, n.iter = Itt, n.thin=1,
  n.update= 10, DIC=T)

QGCFMFSFSQ.mcmc <- as.mcmc(QGCFMFSFSQ)
traceplot(QGCFMFSFSQ.mcmc)
autocorr.plot(QGCFMFSFSQ.mcmc)

#### QGC Random Intercept, Fixed Slope and Quadratic Term ####
initsQGCRMFSFSQ <- function(){
  list(mean.alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100), eta=rnorm(1, mean=1000, sd=100))

  list(mean.alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100), eta=rnorm(1, mean=1000, sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

QGCRMFSFSQ <- jags(data = list("DV","T","N","Occasion"),
  initsQGCRMFSFSQ, model.file="[Path To File]",
  parameters.to.save= c("mean.alpha","var.y",
```



```
  "beta", "eta"), n.chains =2, n.iter = Itt,
  n.burnin=Bi, n.thin=1, DIC=T)

QGCRMFSFSQ <- autojags(QGCRMFSFSQ, n.iter = Itt, n.thin=1,
  n.update= 10, DIC=T)

QGCRMFSFSQ.mcmc <- as.mcmc(QGCRMFSFSQ)
traceplot(QGCRMFSFSQ.mcmc)
autocorr.plot(QGCRMFSFSQ.mcmc)

#### QGC Random Slope, Fixed Intercept and Quadratic Term ####
initsQGCFMRSFSQ <- function(){
  list(alpha=rnorm(1, mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100), eta=rnorm(1, mean=1000, sd=100))

  list(alpha=rnorm(1, mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100), eta=rnorm(1, mean=1000, sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

QGCFMRSFSQ <- jags(data = list("DV","T","N","Occasion"),
  initsQGCFMRSFSQ, model.file="[Path To File]",
  parameters.to.save= c("alpha","var.y",
  "mean.beta", "eta"), n.chains =2, n.iter = Itt,
  n.burnin=Bi, n.thin=1, DIC=T)

QGCFMRSFSQ <- autojags(QGCFMRSFSQ, n.iter = Itt, n.thin=1,
  n.update= 10, DIC=T)

QGCFMRSFSQ.mcmc <- as.mcmc(QGCFMRSFSQ)
traceplot(QGCFMRSFSQ.mcmc)
autocorr.plot(QGCFMRSFSQ.mcmc)
```

```
#### QGC Random Quadratic Term, Fixed Slope and Intercept ####
initsQGCFMFSRSQ <- function(){
  list(alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100), mean.eta=rnorm(1, mean=1000, sd=100))

  list(alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100), mean.eta=rnorm(1, mean=1000, sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

QGCFMFSRSQ <- jags(data = list("DV","T","N","Occasion"),
  initsQGCFMFSRSQ, model.file="[Path To File]",
  parameters.to.save= c("alpha","var.y", "beta",
  "mean.eta"), n.chains =2, n.iter = Itt,
  n.burnin=Bi, n.thin=1, DIC=T)

QGCFMFSRSQ <- autojags(QGCFMFSRSQ, n.iter = Itt, n.thin=1,
  n.update= 10, DIC=T)

QGCFMFSRSQ.mcmc <- as.mcmc(QGCFMFSRSQ)
traceplot(QGCFMFSRSQ.mcmc)
autocorr.plot(QGCFMFSRSQ.mcmc)

#### QGC Random Intercept and Slope, Fixed Quadratic Term ####
initsQGCRMRSFSQ <- function(){
  list(mean.alpha=rnorm(1, mean=1000, sd=100),
  mean.beta=rnorm(1, mean=1000, sd=100), eta=rnorm(1,
  mean=1000, sd=100))

  list(mean.alpha=rnorm(1, mean=1000, sd=100),
  mean.beta=rnorm(1, mean=1000, sd=100), eta=rnorm(1,
```



```
  mean=1000, sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

QGCRMRSFSQ <- jags(data = list("DV","T","N","Occasion"),
  initsQGCRMRSFSQ, model.file="[Path To File]",
  parameters.to.save= c("mean.alpha","var.y",
  "mean.beta", "eta"), n.chains =2, n.iter = Itt,
  n.burnin=Bi, n.thin=1, DIC=T)

QGCRMRSFSQ <- autojags(QGCRMRSFSQ, n.iter = Itt, n.thin=1,
  n.update= 10, DIC=T)

QGCRMRSFSQ.mcmc <- as.mcmc(QGCRMRSFSQ)
traceplot(QGCRMRSFSQ.mcmc)
autocorr.plot(QGCRMRSFSQ.mcmc)

#### QGC Random Intercept and Quadratic Term, Fixed Slope ####
initsQGCRMFSRSQ <- function(){
  list(mean.alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100), mean.eta=rnorm(1, mean=1000, sd=100))

  list(mean.alpha=rnorm(1, mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100), mean.eta=rnorm(1, mean=1000, sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]
```

```
QGCRMFSRSQ <- jags(data = list("DV","T","N","Occasion"),
  initsQGCRMFSRSQ, model.file="[Path To File]",
  parameters.to.save= c("mean.alpha","var.y",
  "beta", "mean.eta"), n.chains =2, n.iter = Itt,
  n.burnin=Bi, n.thin=1, DIC=T)

QGCRMFSRSQ <- autojags(QGCRMFSRSQ, n.iter = Itt, n.thin=1,
  n.update= 10, DIC=T)

QGCRMFSRSQ.mcmc <- as.mcmc(QGCRMFSRSQ)
traceplot(QGCRMFSRSQ.mcmc)
autocorr.plot(QGCRMFSRSQ.mcmc)

#### QGC Random Slope and Quadratic Term, Fixed Intercept ####
initsQGCFMRSRSQ <- function(){
  list(alpha=rnorm(1, mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100), mean.eta=rnorm(1, mean=1000, sd=100))

  list(alpha=rnorm(1, mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100), mean.eta=rnorm(1, mean=1000, sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

QGCFMRSRSQ <- jags(data = list("DV","T","N","Occasion"),
  initsQGCFMRSRSQ, model.file="[Path To File]",
  parameters.to.save= c("alpha","var.y",
  "mean.beta", "mean.eta"), n.chains =2,
  n.iter = Itt, n.burnin=Bi, n.thin=1,
              DIC=T)

QGCFMRSRSQ <- autojags(QGCFMRSRSQ, n.iter = Itt, n.thin=1,
```



```
  n.update= 10, DIC=T)

QGCFMRSRSQ.mcmc <- as.mcmc(QGCFMRSRSQ)
traceplot(QGCFMRSRSQ.mcmc)
autocorr.plot(QGCFMRSRSQ.mcmc)

####### QGC Random Intercept, Slope and Quadratic Term #######
initsQGCRMRSRSQ <- function(){
  list(mean.alpha=rnorm(1, mean=1000, sd=100),
  mean.beta=rnorm(1, mean=1000, sd=100), mean.eta=rnorm(1,
  mean=1000, sd=100))

  list(mean.alpha=rnorm(1, mean=1000, sd=100),
  mean.beta=rnorm(1, mean=1000, sd=100), mean.eta=rnorm(1,
  mean=1000, sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

QGCRMRSRSQ <- jags(data = list("DV","T","N","Occasion"),
  initsQGCRMRSRSQ, model.file="[Path To File]",
  parameters.to.save= c("mean.alpha","var.y",
  "mean.beta", "mean.eta"), n.chains =2,
  n.iter = Itt, n.burnin=Bi, n.thin=1,
              DIC=T)

QGCRMRSRSQ <- autojags(QGCRMRSRSQ, n.iter = Itt, n.thin=1,
  n.update= 10, DIC=T)

QGCRMRSRSQ.mcmc <- as.mcmc(QGCRMRSRSQ)
traceplot(QGCRMRSRSQ.mcmc)
autocorr.plot(QGCRMRSRSQ.mcmc)
```

```
################### ALT Fixed Alpha and Beta #################
initsALTFAFB <- function(){
  list(alpha=rnorm(1,mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100), AR=runif(1,0,1))

  list(alpha=rnorm(1,mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100), AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

ALTFAFB <- jags(data = list("DV","T","N","Occasion"),
   initsALTFAFB, model.file="[Path To File]",
   parameters.to.save= c("intercept","var.y", "slope",
   "AR"), n.chains =2, n.iter = Itt, n.burnin=Bi,
   n.thin=1, DIC=T)

ALTFAFB <- autojags(ALTFAFB, n.iter = Itt, n.thin=1,
   n.update= 10, DIC=T)


ALTFAFB.mcmc <- as.mcmc(ALTFAFB)
autocorr.plot(ALTFAFB.mcmc)
traceplot(ALTFAFB.mcmc)

################# ALT Fixed Alpha, Random Beta ###############
initsALTRAFB <- function(){
  list(mean.alpha=rnorm(1,mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100), AR=runif(1,0,1))

  list(mean.alpha=rnorm(1,mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100), AR=runif(1,0,1))
}
```



```
DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

ALTRAFB <- jags(data = list("DV","T","N","Occasion"),
   initsALTRAFB, model.file="[Path To File]",
   parameters.to.save= c("mean.intercept","var.y",
   "slope","AR"), n.chains =2, n.iter = Itt,
   n.burnin=Bi, n.thin=1, DIC=T)

ALTRAFB <- autojags(ALTRAFB, n.iter = Itt, n.thin=1,
   n.update= 10, DIC=T)


ALTRAFB.mcmc <- as.mcmc(ALTRAFB)
autocorr.plot(ALTRAFB.mcmc)
traceplot(ALTRAFB.mcmc)

################# ALT Random Alpha, Fixed Beta ###############
initsALTFARB <- function(){
  list(alpha=rnorm(1,mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100), AR=runif(1,0,1))

  list(alpha=rnorm(1,mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100), AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

ALTFARB <- jags(data = list("DV","T","N","Occasion"),
   initsALTFARB, model.file="[Path To File]",
   parameters.to.save=c("mean.intercept","var.y",
```

```
   "mean.slope","AR"),n.chains =2, n.iter = Itt,
   n.burnin=Bi, n.thin=1, DIC=T)

ALTFARB <- autojags(ALTFARB, n.iter = Itt, n.thin=1,
   n.update= 10, DIC=T)


ALTFARB.mcmc <- as.mcmc(ALTFARB)
autocorr.plot(ALTFARB.mcmc)
traceplot(ALTFARB.mcmc)

################### ALT Random Alpha and Beta ################
initsALTRARB <- function(){
  list(mean.alpha=rnorm(1,mean=1000, sd=100), mean.beta=rnorm(1
  ,mean=1000, sd=100), AR=runif(1,0,1))

  list(mean.alpha=rnorm(1,mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100), AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

ALTRARB <- jags(data = list("DV","T","N","Occasion"),
   initsALTRARB, model.file="[Path To File]",
   parameters.to.save= c("mean.intercept","var.y",
   "mean.slope","AR"), n.chains =2, n.iter = Itt,
   n.burnin=Bi, n.thin=1, DIC=T)

ALTRARB <- autojags(ALTRARB, n.iter = Itt, n.thin=1,
   n.update= 10, DIC=T)


ALTRARB.mcmc <- as.mcmc(ALTRARB)
autocorr.plot(ALTRARB.mcmc)
```



```
traceplot(ALTRARB.mcmc)

############### ALT Fixed Intercept and Slope ###############
initsALTFMFS <- function(){
  list(alpha=rnorm(1,mean=1000, sd=100), beta=rnorm(1,mean=
  1000, sd=100), AR=runif(1,0,1))

  list(alpha=rnorm(1,mean=1000, sd=100), beta=rnorm(1,mean=
  1000, sd=100), AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

ALTFMFS <- jags(data = list("DV","T","N","Occasion"),
   initsALTFMFS, model.file="[Path To File]",
   parameters.to.save= c("alpha","var.y", "beta","AR"),
           n.chains =2, n.iter = Itt, n.burnin=Bi, n.thin=1,
           DIC=T)

ALTFMFS <- autojags(ALTFMFS, n.iter = Itt, n.thin=1,
   n.update= 10, DIC=T)


ALTFMFS.mcmc <- as.mcmc(ALTFMFS)
autocorr.plot(ALTFMFS.mcmc)
traceplot(ALTFMFS.mcmc)

########## ALT Random Intercept, Fixed and Slope #############
initsALTRMFS <- function(){
  list(mean.alpha=rnorm(1,mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100), AR=runif(1,0,1))

  list(mean.alpha=rnorm(1,mean=1000, sd=100), beta=rnorm(1,
  mean=1000, sd=100), AR=runif(1,0,1))
```

```
}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

ALTRMFS <- jags(data = list("DV","T","N","Occasion"),
   initsALTRMFS, model.file="[Path To File]",
   parameters.to.save= c("mean.alpha","var.y", "beta",
   "AR"), n.chains =2, n.iter = Itt, n.burnin=Bi,
   n.thin=1, DIC=T)

ALTRMFS <- autojags(ALTRMFS, n.iter = Itt, n.thin=1,
   n.update= 10, DIC=T)


ALTRMFS.mcmc <- as.mcmc(ALTRMFS)
autocorr.plot(ALTRMFS.mcmc)
traceplot(ALTRMFS.mcmc)

########## ALT Fixed Intercept, Random and Slope #############
initsALTFMRS <- function(){
  list(alpha=rnorm(1,mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100), AR=runif(1,0,1))

  list(alpha=rnorm(1,mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100), AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

ALTFMRS <- jags(data = list("DV","T","N","Occasion"),
```



```
   initsALTFMRS, model.file="[Path To File]",
   parameters.to.save= c("alpha","var.y", "mean.beta",
   "AR"), n.chains =2, n.iter = Itt, n.burnin=Bi,
   n.thin=1, DIC=T)

ALTFMRS <- autojags(ALTFMRS, n.iter = Itt, n.thin=1,
   n.update= 10, DIC=T)


ALTFMRS.mcmc <- as.mcmc(ALTFMRS)
autocorr.plot(ALTFMRS.mcmc)
traceplot(ALTFMRS.mcmc)

############## ALT Random Intercept and Slope ###############
initsALTRMRS <- function(){
  list(mean.alpha=rnorm(1,mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100), AR=runif(1,0,1))

  list(mean.alpha=rnorm(1,mean=1000, sd=100), mean.beta=rnorm(1,
  mean=1000, sd=100), AR=runif(1,0,1))
}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]

ALTRMRS <- jags(data = list("DV","T","N","Occasion"),
   initsALTRMRS, model.file="[Path To File]",
   parameters.to.save= c("mean.alpha","var.y",
   "mean.beta","AR"),n.chains =2, n.iter = Itt,
   n.burnin=Bi, n.thin=1, DIC=T)

ALTRMRS <- autojags(ALTRMRS, n.iter = Itt, n.thin=1,
   n.update= 10, DIC=T)
```

```
ALTRMRS.mcmc <- as.mcmc(ALTRMRS)
autocorr.plot(ALTRMRS.mcmc)
traceplot(ALTRMRS.mcmc)

#### QGC Random Intercept and Slope, Fixed Quadratic Term ####
################ Group as Predictor for Intercept ##############
initsQGCRMRSFSQPredM <- function(){
  list(gamma00=rnorm(1, mean=1000, sd=100), gamma10=rnorm(1,
  mean=1000, sd=100), gamma20=rnorm(1, mean=1000, sd=100),
  mean.beta=rnorm(1, mean=1000, sd=100), eta=rnorm(1, mean=
  1000, sd=100))

  list(gamma00=rnorm(1, mean=1000, sd=100), gamma10=rnorm(1,
  mean=1000, sd=100), gamma20=rnorm(1, mean=1000, sd=100),
  mean.beta=rnorm(1, mean=1000, sd=100), eta=rnorm(1, mean=
  1000, sd=100))

}

DV <- [Columns of the Data Matrix That Contain the DV]
T <- [Number of Repeated Measures]
N <- [Total Sample Size]
Occasion <- [Indicator Variable for the Different Measurement
 Occasions]


QGCRMRSFSQPredM <- jags(data = list("DV","T","N","Occasion",
   "D1", "D2"), initsQGCRMRSFSQPredM,
   model.file="[Path To File]",
                  parameters.to.save=c("gamma00","gamma10",
                  "gamma20", "var.y", "Cov", "mean.beta",
                  "eta"), n.chains =2, n.iter = Itt,
                  n.burnin=Bi, n.thin=1, DIC=T)


QGCRMRSFSQPredM <- autojags(QGCRMRSFSQPredM, n.iter = Itt,
   n.thin=1, n.update= 10, DIC=T)
```



```
QGCRMRSFSQPredM.mcmc <- as.mcmc(QGCRMRSFSQPredM)
autocorr.plot(QGCRMRSFSQPredM.mcmc)
traceplot(QGCRMRSFSQPredM.mcmc)

#### QGC Random Intercept and Slope, Fixed Quadratic Term ####
################ Group as Predictor for Slope ###############
initsQGCRMRSFSQPredS <- function(){
  list(gamma01=rnorm(1, mean=1000, sd=100), gamma11=rnorm(1,
  mean=1000, sd=100), gamma21=rnorm(1, mean=1000, sd=100),
  mean.alpha=rnorm(1, mean=1000, sd=100), eta=rnorm(1, mean=
  1000, sd=100))


  list(gamma01=rnorm(1, mean=1000, sd=100), gamma11=rnorm(1,
  mean=1000, sd=100), gamma21=rnorm(1, mean=1000, sd=100),
  mean.alpha=rnorm(1, mean=1000, sd=100), eta=rnorm(1, mean=
  1000, sd=100))

}

QGCRMRSFSQPredS <- jags(data = list("DV","T","N","Occasion",
   "D1", "D2"), initsQGCRMRSFSQPredS,
   model.file="[Path To File]",
                  parameters.to.save=c("gamma01","gamma11",
                  "gamma21", "var.y", "Cov", "mean.alpha",
                  "eta"), n.chains =2, n.iter = Itt,
                  n.burnin=Bi, n.thin=1, DIC=T)


QGCRMRSFSQPredS <- autojags(QGCRMRSFSQPredS, n.iter = Itt,
   n.thin=1, n.update= 10, DIC=T)


QGCRMRSFSQPredS.mcmc <- as.mcmc(QGCRMRSFSQPredS)
autocorr.plot(QGCRMRSFSQPredS.mcmc)
traceplot(QGCRMRSFSQPredS.mcmc)
```

```
#### QGC Random Intercept and Slope, Fixed Quadratic Term ####
######### Group as Predictor for Intercept and Slope #########
initsQGCRMRSFSQPredMS <- function(){
  list(gamma01=rnorm(1, mean=1000, sd=100), gamma11=rnorm(1,
  mean=1000, sd=100), gamma21=rnorm(1, mean=1000, sd=100),
  gamma00=rnorm(1, mean=1000, sd=100), gamma10=rnorm(1, mean=
  1000, sd=100), gamma20=rnorm(1, mean=1000, sd=100),
  eta=rnorm(1, mean=1000, sd=100))

  list(gamma01=rnorm(1, mean=1000, sd=100), gamma11=rnorm(1,
  mean=1000, sd=100), gamma21=rnorm(1, mean=1000, sd=100),
  gamma00=rnorm(1, mean=1000, sd=100), gamma10=rnorm(1, mean=
  1000, sd=100), gamma20=rnorm(1, mean=1000, sd=100),
  eta=rnorm(1, mean=1000, sd=100))

}

QGCRMRSFSQPredMS <- jags(data = list("DV","T","N","Occasion",
    "D1", "D2"), initsQGCRMRSFSQPredMS,
    model.file="[Path To File]",
                    parameters.to.save=c("gamma00","gamma10",
                    "gamma20", "gamma01","gamma11", "gamma21",
                    "var.y", "Cov", "eta", "PredCon","PredST",
                    "PredCCT", "PredCCT", "CEACST", "CEACCCT",
"CEACCon"), n.chains =2, n.iter = Itt,
                    n.burnin=Bi, n.thin=1, DIC=T)


QGCRMRSFSQPredMS <- autojags(QGCRMRSFSQPredMS, n.iter = Itt,
    n.thin=1, n.update= 10, DIC=T)

QGCRMRSFSQPredMS.mcmc <- as.mcmc(QGCRMRSFSQPredMS)
autocorr.plot(QGCRMRSFSQPredMS.mcmc)
traceplot(QGCRMRSFSQPredMS.mcmc)

#### QGC Random Intercept and Slope, Fixed Quadratic Term ####
######################### Lognormal Data ######################
DV <- ((-1*DV)+(lambda/6))
```



```
DV <- DV + 5
DV <- log(DV)


initsQGCRMRSFSQPredMSlog <- function(){
  list(gamma01=rnorm(1, mean=6.91, sd=4.61), gamma11=rnorm(1,
   mean=6.91, sd=4.61), gamma21=rnorm(1, mean=6.91, sd=4.61),
   gamma00=rnorm(1, mean=6.91, sd=4.61), gamma10=rnorm(1,
   mean=6.91, sd=4.61), gamma20=rnorm(1, mean=6.91, sd=4.61),
   eta=rnorm(1, mean=6.91, sd=4.61))

  list(gamma01=rnorm(1, mean=6.91, sd=4.61), gamma11=rnorm(1,
  mean=6.91, sd=4.61), gamma21=rnorm(1, mean=6.91, sd=4.61),
  gamma00=rnorm(1, mean=6.91, sd=4.61), gamma10=rnorm(1,
  mean=6.91, sd=4.61), gamma20=rnorm(1, mean=6.91, sd=4.61),
  eta=rnorm(1, mean=6.91, sd=4.61))
  }

QGCRMRSFSQPredMSlog <- jags(data = list("DV","T","N","Occasion",
   "D1", "D2"), initsQGCRMRSFSQPredMSlog,
   model.file="[Path To File]",
                      parameters.to.save=c("gamma00","gamma10",
                      "gamma20", "gamma01","gamma11",
                      "gamma21", "var.y", "Cov", "eta", "PredCon", "PredST",
   "CEACST", "CEACCCT", "CEACCon"),
                      n.chains =2, n.iter = Itt, n.burnin=Bi,
                      n.thin=1, DIC=T)

QGCRMRSFSQPredMSlog <- autojags(QGCRMRSFSQPredMSlog,
   n.iter = Itt, n.thin=1, n.update= 10,
   DIC=T)


QGCRMRSFSQPredMSlog.mcmc <- as.mcmc(QGCRMRSFSQPredMSlog)
autocorr.plot(QGCRMRSFSQPredMSlog.mcmc)
traceplot(QGCRMRSFSQPredMSlog.mcmc)

#### QGC Random Intercept and Slope, Fixed Quadratic Term ####
```

```
################## Gamma Distributed Data ###################
DV <- ((-1*DV)+(lambda/6))
DV <- DV + .0000001


initsQGCRMRSFSQPredMSgamma <- function(){

list(gamma01=rnorm(1, 6.91, 4.61), gamma11=rnorm(1, 6.91, 4.61),
gamma21=rnorm(1, 6.91, 4.61), gamma00=rnorm(1, 6.91, 4.61),
gamma10=rnorm(1, 6.91, 4.61), gamma20=rnorm(1, 6.91, 4.61),
eta=rnorm(1, 6.91, 4.61), shape=runif(1, 0, 20))

list(gamma01=rnorm(1, 6.91, 4.61), gamma11=rnorm(1, 6.91, 4.61),
gamma21=rnorm(1, 6.91, 4.61), gamma00=rnorm(1, 6.91, 4.61),
gamma10=rnorm(1, 6.91, 4.61), gamma20=rnorm(1, 6.91, 4.61),
eta=rnorm(1, 6.91, 4.61), shape=runif(1, 0, 20))

}


QGCRMRSFSQPredMSgamma <- jags(data = list("DV","T","N","Occasion",
                        "D1", "D2"),
                        initsQGCRMRSFSQPredMSgamma,
                        model.file="[Path To File]",
                        parameters.to.save=c("gamma00","gamma10",
                        "gamma20", "gamma01","gamma11", "gamma21",
                        "Cov", "eta", "shape", "PredCon", "PredST", "PredCCT",
   "CEACST", "CEACCCT", "CEACCon"),
   n.chains =2, n.iter = Itt, n.burnin=Bi, n.thin=1, DIC=T)

QGCRMRSFSQPredMSgamma <- autojags(QGCRMRSFSQPredMSgamma,
 n.iter = Itt, n.thin=1, n.update= 10,
 DIC=T)


QGCRMRSFSQPredMSgamma.mcmc <- as.mcmc(QGCRMRSFSQPredMSgamma)
autocorr.plot(QGCRMRSFSQPredMSgamma.mcmc)
traceplot(QGCRMRSFSQPredMSgamma.mcmc)
```


10 Appendix B: JAGS Models

10.1 AR(1) Fixed Mean and AR

```
model{

for (i in 1:N) {

DV[i,1] ~dnorm(mu, tau1)

for(t in 2:T){
  DV[i,t] ~dnorm(Mean[i,t] , tau2)
  Mean[i,t] <- cnst + AR*DV[i,t-1]
}
}

AR ~ dnorm(0, .000000000001)
mu ~ dnorm(0, .000000000001)
cnst <- mu*(1-AR)

tau2 ~ dgamma (0.000001, 0.000001)
tau1 <- (1-(pow(AR,2)))*tau2

var.y <- 1/tau2
}
```

10.2 AR(1) Random Mean and Fixed AR

```
model{

for (i in 1:N) {

mu[i] ~ dnorm(mean.mu, tau.mu)
cnst[i] <- mu[i]*(1-AR)
DV[i,1] ~ dnorm(mu[i], tau1)

for(t in 2:T){
```

```
  DV[i,t] ~dnorm(Mean[i,t] , tau2)
  Mean[i,t] <- cnst[i] + AR*DV[i,t-1]

}
}

AR ~ dnorm(0, .000000000001)
mean.mu ~ dnorm(0, .000000000001)

tau2 ~ dgamma (0.000001, 0.000001)
tau1 <- (1-(pow(AR,2)))*tau2
tau.mu ~ dgamma (0.000001, 0.000001)

var.y <- 1/tau2
}
```

10.3 AR(1) Fixed Mean and Random AR

```
model{

for (i in 1:N) {

AR[i] ~ dnorm(mean.AR, tau.AR)
cnst[i] <- mu*(1-AR[i])
tau1[i] <- (1-(pow(AR[i],2)))*tau2

DV[i,1] ~dnorm(mu, tau1[i])

for(t in 2:T){
  DV[i,t] ~dnorm(Mean[i,t] , tau2)
  Mean[i,t] <- cnst[i] + AR[i]*DV[i,t-1]
}
}

mu ~ dnorm(0, .000000000001)
tau2 ~ dgamma (0.000001, 0.000001)
```



```
mean.AR ~ dnorm(0, .000000000001)
tau.AR ~ dgamma (0.000001, 0.000001)

var.y <- 1/tau2
}
```

10.4 AR(1) Random Mean and Random AR

```
model{

for (i in 1:N) {

LS[i,1:2]~dmnorm(MU[1:2], covmat[1:2,1:2])
mu[i] <-LS[i,1]
AR[i] <- LS[i,2]

cnst[i] <- mu[i]*(1-AR[i])
tau1[i] <- (1-(pow(AR[i],2)))*tau2

DV[i,1] ~ dnorm(mu[i], tau1[i])

for(t in 2:T){
  DV[i,t] ~dnorm(Mean[i,t] , tau2)
  Mean[i,t] <- cnst[i] + AR[i]*DV[i,t-1]

}
}

mean.mu ~ dnorm(0, .0000000001)
mean.AR ~ dnorm(0, .0000000001)
MU[1]<- mean.mu
MU[2] <- mean.AR

tau2 ~ dgamma (0.000001, 0.000001)

covmat[1:2,1:2] ~ dwish(R[,],2)
R[1,1] <- 1000
```

```
R[2,2] <- .01
R[1,2] <- 0
R[2,1]<- R[1,2]
Cov[1:2,1:2] <- inverse(covmat[,])

var.y <- 1/tau2
}
```

10.5 LGC Fixed Intercept and Slope

```
model{
for (i in 1:N) {

for(t in 1:T){

DV[i,t] ~dnorm(mu[i,t] , tau2)
mu[i,t] <- alpha + beta*(Occasion[t]-1)

}
}

alpha ~ dnorm(0, .0000000001)
beta ~ dnorm(0, .0000000001)

tau2 ~ dgamma (0.000001, 0.000001)
var.y <- 1/tau2
}
```

10.6 LGC Random Intercept, Fixed Slope

```
model{

for (i in 1:N) {

alpha[i] ~ dnorm(mean.alpha, tau.alpha)
```



```
for(t in 1:T){

DV[i,t] ~ dnorm(mu[i,t] , tau2)
mu[i,t] <- alpha[i] + beta*(Occasion[t]-1)

}
}

beta ~ dnorm(0, .0000000001)
mean.alpha ~ dnorm(0, .0000000001)

tau2 ~ dgamma (0.000001, 0.000001)
tau.alpha ~ dgamma (0.000001, 0.000001)
var.y <- 1/tau2
}
```

10.7 LGC Fixed Intercept, Random Slope

```
model{

for (i in 1:N) {

beta[i] ~ dnorm(mean.beta, tau.beta)

for(t in 1:T){

DV[i,t] ~ dnorm(mu[i,t] , tau2)
mu[i,t] <- alpha + beta[i]*(Occasion[t]-1)

}
}

alpha ~ dnorm(0, .0000000001)
mean.beta ~ dnorm(0, .0000000001)

tau2 ~ dgamma (0.000001, 0.000001)
tau.beta ~ dgamma (0.000001, 0.000001)
```

```
var.y <- 1/tau2
}
```

10.8 LGC Random Intercept and Slope

```
model{

for (i in 1:N) {

LS[i,1:2]~dmnorm(MU[1:2], covmat[1:2,1:2])
alpha[i] <-LS[i,1]
beta[i] <- LS[i,2]

for(t in 1:T){

DV[i,t] ~ dnorm(mu[i,t] , tau2)
mu[i,t] <- alpha[i] + beta[i]*(Occasion[t]-1)

}
}

mean.alpha ~ dnorm(0, .0000000001)
mean.beta ~ dnorm(0, .0000000001)
MU[1]<- mean.alpha
MU[2] <- mean.beta

tau2 ~ dgamma (0.000001, 0.000001)
covmat[1:2,1:2] ~ dwish(R[,],2)

R[1,1] <- 1000
R[2,2] <- 500
R[1,2] <- 0
R[2,1]<- R[1,2]
Cov[1:2,1:2] <- inverse(covmat[,])
rho1 <- Cov[1,2]/sqrt(Cov[1,1]*Cov[2,2])

var.y <- 1/tau2
```



```
}
```

10.9 QGC Fixed Intercept, Slope, and Quadratic Term

```
model{

for (i in 1:N) {

for(t in 1:T){

DV[i,t] ~dnorm(mu[i,t] , tau2)
mu[i,t] <- alpha + beta*(Occasion[t]-1) +
   eta*pow((Occasion[t]-1),2)

}
}

alpha ~ dnorm(0, .0000000001)
beta ~ dnorm(0, .0000000001)
eta ~ dnorm(0, .0000000001)

tau2 ~ dgamma (0.000001, 0.000001)
var.y <- 1/tau2

}
```

10.10 QGC Random Intercept, Fixed Slope and Quadratic Term

```
model{

for (i in 1:N) {

alpha[i] ~ dnorm(mean.alpha, tau.alpha)

for(t in 1:T){
```

```
DV[i,t] ~dnorm(mu[i,t] , tau2)
mu[i,t] <- alpha[i] + beta*(Occasion[t]-1) +
   eta*pow((Occasion[t]-1),2)

}
}

mean.alpha ~ dnorm(0, .0000000001)
beta ~ dnorm(0, .0000000001)
eta ~ dnorm(0, .0000000001)

tau2 ~ dgamma (0.000001, 0.000001)
tau.alpha ~ dgamma (0.000001, 0.000001)
var.y <- 1/tau2
}
```

10.11 QGC Random Slope, Fixed Intercept and Quadratic Term

```
model{
for (i in 1:N) {

beta[i] ~ dnorm(mean.beta, tau.beta)

for(t in 1:T){
DV[i,t] ~dnorm(mu[i,t] , tau2)
mu[i,t] <- alpha + beta[i]*(Occasion[t]-1) +
   eta*pow((Occasion[t]-1),2)

}
}

alpha ~ dnorm(0, .0000000001)
mean.beta ~ dnorm(0, .0000000001)
eta ~ dnorm(0, .0000000001)
```



```
tau2 ~ dgamma (0.000001, 0.000001)
tau.beta ~ dgamma (0.000001, 0.000001)
var.y <- 1/tau2
}
```

10.12 QGC Random Quadratic Term, Fixed Intercept and Slope

```
model{

for (i in 1:N) {

eta[i] ~ dnorm(mean.eta, tau.eta)

for(t in 1:T){
DV[i,t] ~dnorm(mu[i,t] , tau2)
mu[i,t] <- alpha + beta*(Occasion[t]-1) +
   eta[i]*pow((Occasion[t]-1),2)

}
}

mean.eta ~ dnorm(0, .0000000001)
beta ~ dnorm(0, .0000000001)
alpha ~ dnorm(0, .0000000001)

tau2 ~ dgamma (0.000001, 0.000001)
tau.eta ~ dgamma (0.000001, 0.000001)
var.y <- 1/tau2
}
```

10.13 QGC Random Intercept and Slope, Fixed Quadratic Term

```
model{
```

```
for (i in 1:N) {

LS[i,1:2]~dmnorm(MU[1:2], covmat[1:2,1:2])
alpha[i] <-LS[i,1]
beta[i] <- LS[i,2]

for(t in 1:T){
DV[i,t] ~dnorm(mu[i,t] , tau2)
mu[i,t] <- alpha[i] + beta[i]*(Occasion[t]-1) +
   eta*pow((Occasion[t]-1),2)

}
}

mean.alpha ~ dnorm(0, .0000000001)
mean.beta ~ dnorm(0, .0000000001)
MU[1]<- mean.alpha
MU[2] <- mean.beta
eta ~ dnorm(0, .0000000001)

tau2 ~ dgamma (0.000001, 0.000001)
covmat[1:2,1:2] ~ dwish(R[,],2)

R[1,1] <- 1000
R[2,2] <- 500
R[1,2] <- 0
R[2,1]<- R[1,2]
Cov[1:2,1:2] <- inverse(covmat[,])
rho1 <- Cov[1,2]/sqrt(Cov[1,1]*Cov[2,2])
var.y <- 1/tau2
}
```


10.14 QGC Random Intercept and Quadratic Term, Fixed Slope

```
model{

for (i in 1:N) {

LS[i,1:2]~dmnorm(MU[1:2], covmat[1:2,1:2])
alpha[i] <-LS[i,1]
eta[i] <-LS[i,2]

for(t in 1:T){
DV[i,t] ~dnorm(mu[i,t] , tau2)
mu[i,t] <- alpha[i] + beta*(Occasion[t]-1) +
   eta[i]*pow((Occasion[t]-1),2)

}
}

mean.alpha ~ dnorm(0, .0000000001)
mean.eta ~ dnorm(0, .0000000001)
MU[1]<- mean.alpha
MU[2] <- mean.eta
beta ~ dnorm(0, .0000000001)

tau2 ~ dgamma (0.000001, 0.000001)
covmat[1:2,1:2] ~ dwish(R[,],2)
R[1,1] <- 1000
R[2,2] <- 500
R[1,2] <- 0
R[2,1]<- R[1,2]
Cov[1:2,1:2] <- inverse(covmat[,])
rho1 <- Cov[1,2]/sqrt(Cov[1,1]*Cov[2,2])
var.y <- 1/tau2
}
```

10.15 QGC Random Slope and Quadratic Term, Fixed Intercept

```
model{

for (i in 1:N) {
LS[i,1:2]~dmnorm(MU[1:2], covmat[1:2,1:2])
beta[i] <-LS[i,1]
eta[i] <- LS[i,2]

for(t in 1:T){
DV[i,t] ~dnorm(mu[i,t] , tau2)
mu[i,t] <- alpha + beta[i]*(Occasion[t]-1) +
   eta[i]*pow((Occasion[t]-1),2)

}
}

mean.beta ~ dnorm(0, .0000000001)
mean.eta ~ dnorm(0, .0000000001)
MU[1]<- mean.beta
MU[2] <- mean.eta
alpha ~ dnorm(0, .0000000001)

tau2 ~ dgamma (0.000001, 0.000001)
covmat[1:2,1:2] ~ dwish(R[,],2)
R[1,1] <- 1000
R[2,2] <- 500
R[1,2] <- 0
R[2,1]<- R[1,2]
Cov[1:2,1:2] <- inverse(covmat[,])
rho1 <- Cov[1,2]/sqrt(Cov[1,1]*Cov[2,2])
var.y <- 1/tau2
}
```


10.16 QGC Random Intercept, Slope, and Quadratic Term

```
model{

for (i in 1:N) {

LS[i,1:3]~dmnorm(MU[1:3], covmat[1:3,1:3])
alpha[i] <-LS[i,1]
beta[i] <- LS[i,2]
eta[i] <- LS[i,3]

for(t in 1:T){
DV[i,t] ~dnorm(mu[i,t] , tau2)
mu[i,t] <- alpha[i] + beta[i]*(Occasion[t]-1) +
   eta[i]*pow((Occasion[t]-1),2)

}
}

mean.alpha ~ dnorm(0, .0000000001)
mean.beta ~ dnorm(0, .0000000001)
mean.eta ~ dnorm(0, .0000000001)
MU[1]<- mean.alpha
MU[2] <- mean.beta
MU[3] <- mean.eta

tau2 ~ dgamma (0.000001, 0.000001)
covmat[1:3,1:3] ~ dwish(R[,],3)
R[1,1] <- 1000
R[2,2] <- 500
R[3,3] <- 100 #500
R[1,2] <- 0
R[1,3] <- 0
R[2,3] <- 0
R[2,1]<- R[1,2]
R[3,1]<- R[1,3]
R[3,2]<- R[2,3]
```

```
Cov[1:3,1:3] <- inverse(covmat[,])
rho1 <- Cov[1,2]/sqrt(Cov[1,1]*Cov[2,2])
rho2 <- Cov[1,3]/sqrt(Cov[1,1]*Cov[3,3])
rho3 <- Cov[2,3]/sqrt(Cov[2,2]*Cov[3,3])
var.y <- 1/tau2
}
```

10.17 ALT Fixed Alpha and Beta

```
model{

for (i in 1:N) {

DV[i,1] ~dnorm(mu[i,1] , tau1)
mu[i,1] <- (1/(1-AR))*alpha - (AR/pow((1-AR),2))*beta

for(t in 2:T){
  DV[i,t] ~dnorm(mu[i,t] , tau2)
  mu[i,t] <- alpha + beta*Occasion[t-1] + AR*DV[i,t-1]

}
}

AR ~ dnorm(0, .0000000001)
alpha ~ dnorm(0, .000000000001)
beta ~ dnorm(0, .000000000001)

tau1 <-(1-pow(AR,2))*tau2
tau2 ~ dgamma (0.000001, 0.000001)

var.y <- 1/tau2
var.y1 <- 1/tau1

intercept <- ((alpha)/(1-AR)) - ((beta*AR)/pow((1-AR),2))
slope <- beta/(1-AR)
}
```


10.18 ALT Random Alpha, Fixed Beta

```
model{

for (i in 1:N) {

alpha[i] ~ dnorm(mean.alpha, tau.alpha)
intercept[i] <- ((alpha[i])/(1-AR)) - ((beta*AR)/pow((1-AR),2))

DV[i,1] ~dnorm(mu[i,1] , tau1)
mu[i,1] <- (1/(1-AR))*alpha[i] - (AR/pow((1-AR),2))*beta

for(t in 2:T){
  DV[i,t] ~dnorm(mu[i,t] , tau2)
  mu[i,t] <- alpha[i] + beta*Occasion[t-1] + AR*DV[i,t-1]

}



AR ~ dnorm(0, .0000000001)
mean.alpha ~ dnorm(0, .000000000001)
beta ~ dnorm(0, .000000000001)

tau1 <-(1-pow(AR,2))*tau2
tau2 ~ dgamma (0.000001, 0.000001)
tau.alpha ~ dgamma (0.000001, 0.000001)

var.y <- 1/tau2
var.y1 <- 1/tau1

mean.intercept <- mean(intercept[])
slope <- beta/(1-AR)
}
```

10.19 ALT Fixed Alpha, Random Beta

```
model{

for (i in 1:N) {

beta[i] ~ dnorm(mean.beta, tau.beta)
intercept[i] <- ((alpha)/(1-AR)) - ((beta[i]*AR)/pow((1-AR),2))
slope[i] <- beta[i]/(1-AR)

DV[i,1] ~dnorm(mu[i,1] , tau1)
mu[i,1] <- (1/(1-AR))*alpha - (AR/pow((1-AR),2))*beta[i]

for(t in 2:T){
  DV[i,t] ~dnorm(mu[i,t] , tau2)
  mu[i,t] <- alpha + beta[i]*Occasion[t-1] + AR*DV[i,t-1]

}
}


AR ~ dnorm(0, .0000000001)
alpha ~ dnorm(0, .000000000001)
mean.beta ~ dnorm(0, .000000000001)

tau1 <-(1-pow(AR,2))*tau2
tau2 ~ dgamma (0.000001, 0.000001)
tau.beta ~ dgamma (0.000001, 0.000001)

var.y <- 1/tau2
var.y1 <- 1/tau1

mean.intercept <- mean(intercept[])
mean.slope <- mean(slope[])
}
```


10.20 ALT Random Alpha and Beta

```
model{

for (i in 1:N) {

LS[i,1:2]~dmnorm(MU[1:2], covmat[1:2,1:2])
alpha[i] <-LS[i,1]
beta[i] <- LS[i,2]
intercept[i] <- ((alpha[i])/(1-AR)) - ((beta[i]*AR)/pow((1-AR),2))
slope[i] <- beta[i]/(1-AR)

DV[i,1] ~dnorm(mu[i,1] , tau1)
mu[i,1] <- (1/(1-AR))*alpha[i] - (AR/pow((1-AR),2))*beta[i]

for(t in 2:T){
  DV[i,t] ~dnorm(mu[i,t] , tau2)
  mu[i,t] <- alpha[i] + beta[i]*Occasion[t-1] + AR*DV[i,t-1]

}
}


AR ~ dnorm(0, .0000000001)
mean.alpha ~ dnorm(0, .0000000001)
mean.beta ~ dnorm(0, .0000000001)
MU[1]<- mean.alpha
MU[2] <- mean.beta


tau1 <-(1-pow(AR,2))*tau2
tau2 ~ dgamma (0.000001, 0.000001)

covmat[1:2,1:2] ~ dwish(R[,],2)
R[1,1] <- 1000
R[2,2] <- 500
R[1,2] <- 0
R[2,1]<- R[1,2]
```

```
Cov[1:2,1:2] <- inverse(covmat[,])
rho1 <- Cov[1,2]/sqrt(Cov[1,1]*Cov[2,2])

var.y <- 1/tau2
var.y1 <- 1/tau1

mean.intercept <- mean(intercept[])
mean.slope <- mean(slope[])
}
```

10.21 ALT Fixed Intercept and Slope

```
model{

for (i in 1:N) {

DV[i,1] ~ dnorm(trend[i,1] , tau1)
trend[i,1] <-  alpha

for(t in 2:T){
DV[i,t] ~dnorm(mu[i,t] , tau2)
trend[i,t] <- alpha + beta*Occasion[t-1]
mu[i,t] <- trend[i,t] + AR*(DV[i,t-1] - trend[i,t-1])
}
}

AR ~ dnorm(0, .0000000001)
alpha ~ dnorm(0, .000000000001)
beta ~ dnorm(0, .000000000001)
tau1 <-(1-pow(AR,2))*tau2
tau2 ~ dgamma (0.000001, 0.000001)

var.y <- 1/tau2
var.y1 <- 1/tau1
}
```


10.22 ALT Random Intercept, Fixed Slope

```
model{

for (i in 1:N) {

alpha[i] ~ dnorm(mean.alpha, tau.alpha)
DV[i,1] ~ dnorm(trend[i,1] , tau1)
trend[i,1] <-  alpha[i]

for(t in 2:T){
DV[i,t] ~dnorm(mu[i,t] , tau2)
trend[i,t] <- alpha[i] + beta*Occasion[t-1]
mu[i,t] <- trend[i,t] + AR*(DV[i,t-1] - trend[i,t-1])
}
}

AR ~ dnorm(0, .0000000001)
mean.alpha ~ dnorm(0, .000000000001)
beta ~ dnorm(0, .000000000001)

tau1 <-(1-pow(AR,2))*tau2
tau2 ~ dgamma (0.000001, 0.000001)
tau.alpha ~ dgamma (0.000001, 0.000001)

var.y <- 1/tau2
var.y1 <- 1/tau1
}
```

10.23 ALT Fixed Intercept, Random Slope

```
model{

for (i in 1:N) {

beta[i] ~ dnorm(mean.beta, tau.beta)
DV[i,1] ~ dnorm(trend[i,1] , tau1)
```

```
trend[i,1] <-  alpha

for(t in 2:T){
DV[i,t] ~dnorm(mu[i,t] , tau2)
trend[i,t] <- alpha + beta[i]*Occasion[t-1]
mu[i,t] <- trend[i,t] + AR*(DV[i,t-1] - trend[i,t-1])
}
}

AR ~ dnorm(0, .0000000001)
alpha ~ dnorm(0, .000000000001)
mean.beta ~ dnorm(0, .000000000001)

tau1 <-(1-pow(AR,2))*tau2
tau2 ~ dgamma (0.000001, 0.000001)
tau.beta ~ dgamma (0.000001, 0.000001)

var.y <- 1/tau2
var.y1 <- 1/tau1
}
```

10.24 ALT Random Intercept and Slope

```
model{

for (i in 1:N) {

LS[i,1:2]~dmnorm(MU[1:2], covmat[1:2,1:2])
alpha[i] <-LS[i,1]
beta[i] <- LS[i,2]

DV[i,1] ~ dnorm(trend[i,1] , tau1)
trend[i,1] <-  alpha[i]

for(t in 2:T){
DV[i,t] ~dnorm(mu[i,t] , tau2)
trend[i,t] <- alpha[i] + beta[i]*Occasion[t-1]
```



```
mu[i,t] <- trend[i,t] + AR*(DV[i,t-1] - trend[i,t-1])
}
}

AR ~ dnorm(0, .0000000001)
mean.alpha ~ dnorm(0, .0000000001)
mean.beta ~ dnorm(0, .0000000001)
MU[1]<- mean.alpha
MU[2] <- mean.beta

tau1 <-(1-pow(AR,2))*tau2
tau2 ~ dgamma (0.000001, 0.000001)

covmat[1:2,1:2] ~ dwish(R[,],2)
R[1,1] <- 1000
R[2,2] <- 500
R[1,2] <- 0
R[2,1]<- R[1,2]
Cov[1:2,1:2] <- inverse(covmat[,])
rho1 <- Cov[1,2]/sqrt(Cov[1,1]*Cov[2,2])
var.y <- 1/tau2
var.y1 <- 1/tau1
}
```

10.25 QGC Random Intercept and Slope, Fixed Quadratic Term: Predictor for Intercept

```
model{

for (i in 1:N) {

LS[i,1:2]~dmnorm(MU[i,1:2], covmat[1:2,1:2])
MU[i,1]<- gamma00 + gamma10*D1[i] + gamma20*D2[i]
MU[i,2] <- mean.beta
alpha[i] <-LS[i,1]
beta[i] <- LS[i,2]
```

```
for(t in 1:T){
DV[i,t] ~dnorm(mu[i,t] , tau2)
mu[i,t] <- alpha[i] + beta[i]*(Occasion[t]-1) +
   eta*pow((Occasion[t]-1),2)

}
}

gamma00 ~ dnorm(0, .0000000001)
gamma10 ~ dnorm(0, .0000000001)
gamma20 ~ dnorm(0, .0000000001)

mean.beta ~ dnorm(0, .0000000001)
eta ~ dnorm(0, .0000000001)

tau2 ~ dgamma (0.000001, 0.000001)
covmat[1:2,1:2] ~ dwish(R[,],2)
R[1,1] <- 1000
R[2,2] <- 500
R[1,2] <- 0
R[2,1]<- R[1,2]
Cov[1:2,1:2] <- inverse(covmat[,])
rho1 <- Cov[1,2]/sqrt(Cov[1,1]*Cov[2,2])

var.y <- 1/tau2

}
```

10.26 QGC Random Intercept and Slope, Fixed Quadratic Term: Predictor for Slope

```
model{

for (i in 1:N) {
```



```
LS[i,1:2]~dmnorm(MU[i,1:2], covmat[1:2,1:2])
MU[i,1] <- mean.alpha
MU[i,2] <- gamma01 + gamma11*D1[i] + gamma21*D2[i]
alpha[i] <-LS[i,1]
beta[i] <- LS[i,2]

for(t in 1:T){
DV[i,t] ~dnorm(mu[i,t] , tau2)
mu[i,t] <- alpha[i] + beta[i]*(Occasion[t]-1) +
   eta*pow((Occasion[t]-1),2)

}
}

mean.alpha ~ dnorm(0, .0000000001)
gamma01 ~ dnorm(0, .0000000001)
gamma11 ~ dnorm(0, .0000000001)
gamma21 ~ dnorm(0, .0000000001)
eta ~ dnorm(0, .0000000001)

tau2 ~ dgamma (0.000001, 0.000001)
covmat[1:2,1:2] ~ dwish(R[,],2)
R[1,1] <- 1000
R[2,2] <- 500
R[1,2] <- 0
R[2,1]<- R[1,2]
Cov[1:2,1:2] <- inverse(covmat[,])
rho1 <- Cov[1,2]/sqrt(Cov[1,1]*Cov[2,2])
var.y <- 1/tau2

}
```

10.27 QGC Random Intercept and Slope, Fixed Quadratic Term: Predictor for Intercept and Slope

```
model{
```

```
for (i in 1:N) {

LS[i,1:2]~dmnorm(MU[i,1:2], covmat[1:2,1:2])
MU[i,1] <- gamma00 + gamma10*D1[i] + gamma20*D2[i]
MU[i,2] <- gamma01 + gamma11*D1[i] + gamma21*D2[i]
alpha[i] <-LS[i,1]
beta[i] <- LS[i,2]

for(t in 1:T){
DV[i,t] ~dnorm(mu[i,t] , tau2)
mu[i,t] <- alpha[i] + beta[i]*(Occasion[t]-1) +
   eta*pow((Occasion[t]-1),2)

}
}

gamma00 ~ dnorm(0, .0000000001)
gamma10 ~ dnorm(0, .0000000001)
gamma20 ~ dnorm(0, .0000000001)
gamma01 ~ dnorm(0, .0000000001)
gamma11 ~ dnorm(0, .0000000001)
gamma21 ~ dnorm(0, .0000000001)
eta ~ dnorm(0, .0000000001)

tau2 ~ dgamma (0.000001, 0.000001)
covmat[1:2,1:2] ~ dwish(R[,],2)
R[1,1] <- 1000
R[2,2] <- 500
R[1,2] <- 0
R[2,1]<- R[1,2]
Cov[1:2,1:2] <- inverse(covmat[,])
rho1 <- Cov[1,2]/sqrt(Cov[1,1]*Cov[2,2])
var.y <- 1/tau2

PredCon[1] <- gamma00 + (gamma01*(0)) + (eta*(0))
PredCon[2] <- gamma00 + (gamma01*(1)) + (eta*(1))
PredCon[3] <- gamma00 + (gamma01*(2)) + (eta*(4))
```



```
PredCon[4] <- gamma00 + (gamma01*(3)) + (eta*(9))
PredCon[5] <- gamma00 + (gamma01*(4)) + (eta*(16))
PredCon[6] <- gamma00 + (gamma01*(5)) + (eta*(25))
PredCon[7] <- gamma00 + (gamma01*(6)) + (eta*(36))

PredST[1] <- (gamma00+gamma10) + ((gamma01+gamma11)*(0)) +
 (eta*(0))
PredST[2] <- (gamma00+gamma10) + ((gamma01+gamma11)*(1)) +
 (eta*(1))
PredST[3] <- (gamma00+gamma10) + ((gamma01+gamma11)*(2)) +
 (eta*(4))
PredST[4] <- (gamma00+gamma10) + ((gamma01+gamma11)*(3)) +
 (eta*(9))
PredST[5] <- (gamma00+gamma10) + ((gamma01+gamma11)*(4)) +
 (eta*(16))
PredST[6] <- (gamma00+gamma10) + ((gamma01+gamma11)*(5)) +
 (eta*(25))
PredST[7] <- (gamma00+gamma10) + ((gamma01+gamma11)*(6)) +
 (eta*(36))

PredCCT[1] <- (gamma00+gamma20) + ((gamma01+gamma21)*(0)) +
  (eta*(0))
PredCCT[2] <- (gamma00+gamma20) + ((gamma01+gamma21)*(1)) +
  (eta*(1))
PredCCT[3] <- (gamma00+gamma20) + ((gamma01+gamma21)*(2)) +
      (eta*(4))
PredCCT[4] <- (gamma00+gamma20) + ((gamma01+gamma21)*(3)) +
  (eta*(9))
PredCCT[5] <- (gamma00+gamma20) + ((gamma01+gamma21)*(4)) +
  (eta*(16))
PredCCT[6] <- (gamma00+gamma20) + ((gamma01+gamma21)*(5)) +
  (eta*(25))
PredCCT[7] <- (gamma00+gamma20) + ((gamma01+gamma21)*(6)) +
  (eta*(36))

INBSTvsCon <- (PredST[2] + PredST[3] + PredST[4] + PredST[5] +
PredST[6] + PredST[7]) - (PredCon[2] + PredCon[3] +
PredCon[4] + PredCon[5] + PredCon[6] + PredCon[7])
```

```
INBSTvsCCT <- (PredST[2] + PredST[3] + PredST[4] + PredST[5] +
PredST[6] + PredST[7]) - (PredCCT[2] + PredCCT[3] +
PredCCT[4] + PredCCT[5] + PredCCT[6] + PredCCT[7])
INBCCTvsCon <- (PredCCT[2] + PredCCT[3] + PredCCT[4] + PredCCT[5] +
PredCCT[6] + PredCCT[7]) - (PredCon[2] + PredCon[3] +
PredCon[4] + PredCon[5] + PredCon[6] + PredCon[7])

INMBConvsST <- (PredCon[2] + PredCon[3] + PredCon[4] + PredCon[5] +
PredCon[6] + PredCon[7]) - (PredST[2] + PredST[3] +
PredST[4] + PredST[5] + PredST[6] + PredST[7])
INMBCCTvsST <- (PredCCT[2] + PredCCT[3] + PredCCT[4] + PredCCT[5] +
PredCCT[6] + PredCCT[7]) - (PredST[2] + PredST[3] +
PredST[4] + PredST[5] + PredST[6] + PredST[7])
INMBConvsCCT <- (PredCon[2] + PredCon[3] + PredCon[4] + PredCon[5] +
PredCon[6] + PredCon[7]) - (PredCCT[2] + PredCCT[3] +
PredCCT[4] + PredCCT[5] + PredCCT[6] + PredCCT[7])

CEACST <- step(INMBSTvsCon)*step(INMBSTvsCCT)
CEACCCT <- step(INMBCCTvsST)*step(INMBCCTvsCon)
CEACCon <- step(INMBConvsST)*step(INMBConvsCCT)

}
```

10.28 Three-level QGC Model with Random Intercept and Slope, Fixed Quadratic Term

```
model{

for (j in 1:C) {

mean.alpha[j] ~ dnorm(MA, TauA)
mean.beta[j] ~ dnorm(MB, TauB)

}
```



```
for (i in 1:N) {

MU[i,1] <- mean.alpha[Centre[i]]
MU[i,2] <- mean.beta[Centre[i]]

LS[i,1:2]~dmnorm(MU[i,1:2], covmat[1:2,1:2])
alpha[i] <-LS[i,1]
beta[i] <- LS[i,2]

for(t in 1:T){
DV[i,t] ~dnorm(mu[i,t] , tau2)
 mu[i,t] <- alpha[i] + beta[i]*(Occasion[t]-1) +
      eta*pow((Occasion[t]-1),2)

}
}

MA ~ dnorm(0, .0000000001)
MB ~ dnorm(0, .0000000001)
eta ~ dnorm(0, .0000000001)

tau2 ~ dgamma (0.000001, 0.000001)
TauA ~ dgamma (0.000001, 0.000001)
TauB ~ dgamma (0.000001, 0.000001)

VarA <- 1/TauA
VarB <- 1/TauB

covmat[1:2,1:2] ~ dwish(R[,],2)
R[1,1] <- 1000
R[2,2] <- 500
R[1,2] <- 0
R[2,1]<- R[1,2]

Cov[1:2,1:2] <- inverse(covmat[,])
rho1 <- Cov[1,2]/sqrt(Cov[1,1]*Cov[2,2])
var.y <- 1/tau2
}
```

10.29 Three-level QGC Model with Random Intercept and Slope, Fixed Quadratic Term: Predictor for Intercept and Slope

```
model{



for (j in 1:C) {
# These lines draw separate intercepts and regression
# coefficients for each centre in the study. Thus they
# allow for different treatment means in each centre.

# If one want the initial mean NB of the treatments to be
# the same across centres, the following three lines should
# be removed, and each individual should get a common gamma00,
# gamma10, and gamma20

gamma00[j] ~ dnorm(Mugamma00, Taugamma00)
gamma10[j] ~ dnorm(Mugamma10, Taugamma10)
gamma20[j] ~ dnorm(Mugamma20, Taugamma20)

# If one want the change in NB of the treatments to be the
# same across centres, the following three lines should be
# removed, and each individual should get a common gamma10,
# gamma11, and gamma21

gamma01[j] ~ dnorm(Mugamma01, Taugamma01)
gamma11[j] ~ dnorm(Mugamma11, Taugamma11)
gamma21[j] ~ dnorm(Mugamma21, Taugamma21)
}


for (i in 1:N) {
```



```
LS[i,1:2]~ dmnorm(MU[i,1:2], covmat[1:2,1:2])

MU[i,1] <- gamma00[Centre[i]] + gamma10[Centre[i]]*D1[i] +
           gamma20[Centre[i]]*D2[i]

MU[i,2] <- gamma01[Centre[i]] + gamma11[Centre[i]]*D1[i] +
           gamma21[Centre[i]]*D2[i]

alpha[i] <-LS[i,1]
beta[i] <- LS[i,2]


for(t in 1:T){

DV[i,t] ~ dnorm(mu[i,t] , tau2)
mu[i,t] <- alpha[i] + beta[i]*(Occasion[t]-1) +
   eta*pow((Occasion[t]-1),2)

}
}

Mugamma00 ~ dnorm(0, .0000000001)
Mugamma10 ~ dnorm(0, .0000000001)
Mugamma20 ~ dnorm(0, .0000000001)

Mugamma01 ~ dnorm(0, .0000000001)
Mugamma11 ~ dnorm(0, .0000000001)
Mugamma21 ~ dnorm(0, .0000000001)

eta ~ dnorm(0, .0000000001)

tau2 ~ dgamma (0.000001, 0.000001)

Taugamma00 ~ dgamma (0.000001, 0.000001)
Taugamma10 ~ dgamma (0.000001, 0.000001)
Taugamma20 ~ dgamma (0.000001, 0.000001)

Taugamma01 ~ dgamma (0.000001, 0.000001)
```

```
Taugamma11 ~ dgamma (0.000001, 0.000001)
Taugamma21 ~ dgamma (0.000001, 0.000001)


Vargamma00 <- 1/Taugamma00
Vargamma10 <- 1/Taugamma10
Vargamma20 <- 1/Taugamma20

Vargamma01 <- 1/Taugamma01
Vargamma11 <- 1/Taugamma11
Vargamma21 <- 1/Taugamma21


# Note that we are using the same covariance matrix for
# people form each centre in the study. If differences in
# variance between the centres is likely, we should draw
# separate R matrices the same way we draw separate gamma's
# above.

covmat[1:2,1:2] ~ dwish(R[,],2)
R[1,1] <- 1000
R[2,2] <- 500
R[1,2] <- 0
R[2,1]<- R[1,2]

Cov[1:2,1:2] <- inverse(covmat[,])
rho1 <- Cov[1,2]/sqrt(Cov[1,1]*Cov[2,2])
var.y <- 1/tau2
}
```

10.30 QGC Random Intercept and Slope, Fixed Quadratic Term: Predictor for Intercept and Slope, Log-normal Distributed Data

```
model{

for (i in 1:N) {
```



```
LS[i,1:2]~dmnorm(MU[i,1:2], covmat[1:2,1:2])
MU[i,1] <- gamma00 + gamma10*D1[i] + gamma20*D2[i]
MU[i,2] <- gamma01 + gamma11*D1[i] + gamma21*D2[i]
alpha[i] <-LS[i,1]
beta[i] <- LS[i,2]

for(t in 1:T){
DV[i,t] ~dnorm(mu[i,t] , tau2)
mu[i,t] <- alpha[i] + beta[i]*(Occasion[t]-1) +
   eta*pow((Occasion[t]-1),2)

}
}

gamma00 ~ dnorm(0, .0000000001)
gamma10 ~ dnorm(0, .0000000001)
gamma20 ~ dnorm(0, .0000000001)
gamma01 ~ dnorm(0, .0000000001)
gamma11 ~ dnorm(0, .0000000001)
gamma21 ~ dnorm(0, .0000000001)
eta ~ dnorm(0, .0000000001)

tau2 ~ dgamma (0.000001, 0.000001)
covmat[1:2,1:2] ~ dwish(R[,],2)
R[1,1] <- 3
R[2,2] <- 2
R[1,2] <- 0
R[2,1]<- R[1,2]
Cov[1:2,1:2] <- inverse(covmat[,])
rho1 <- Cov[1,2]/sqrt(Cov[1,1]*Cov[2,2])
var.y <- 1/tau2

PredCon[1] <- exp(gamma00 + (gamma01*(0)) + (eta*(0)) + 1/(tau2*2))
PredCon[2] <- exp(gamma00 + (gamma01*(1)) + (eta*(1)) + 1/(tau2*2))
PredCon[3] <- exp(gamma00 + (gamma01*(2)) + (eta*(4)) + 1/(tau2*2))
PredCon[4] <- exp(gamma00 + (gamma01*(3)) + (eta*(9)) + 1/(tau2*2))
PredCon[5] <- exp(gamma00 + (gamma01*(4)) + (eta*(16))) + 1/(tau2*2))
```

```
PredCon[6] <- exp(gamma00 + (gamma01*(5)) + (eta*(25)) + 1/(tau2*2))
PredCon[7] <- exp(gamma00 + (gamma01*(6)) + (eta*(36))) + 1/(tau2*2))

PredST[1] <- exp((gamma00+gamma10) + ((gamma01+gamma11)*(0)) +
 (eta*(0)) + 1/(tau2*2))
PredST[2] <- exp((gamma00+gamma10) + ((gamma01+gamma11)*(1)) +
 (eta*(1)) + 1/(tau2*2))
PredST[3] <- exp((gamma00+gamma10) + ((gamma01+gamma11)*(2)) +
 (eta*(4)) + 1/(tau2*2))
PredST[4] <- exp((gamma00+gamma10) + ((gamma01+gamma11)*(3)) +
 (eta*(9)) + 1/(tau2*2))
PredST[5] <- exp((gamma00+gamma10) + ((gamma01+gamma11)*(4)) +
 (eta*(16)) + 1/(tau2*2))
PredST[6] <- exp((gamma00+gamma10) + ((gamma01+gamma11)*(5)) +
 (eta*(25)) + 1/(tau2*2))
PredST[7] <- exp((gamma00+gamma10) + ((gamma01+gamma11)*(6)) +
 (eta*(36)) + 1/(tau2*2))

PredCCT[1] <- exp((gamma00+gamma20) + ((gamma01+gamma21)*(0)) +
  (eta*(0)) + 1/(tau2*2))
PredCCT[2] <- exp((gamma00+gamma20) + ((gamma01+gamma21)*(1)) +
  (eta*(1)) + 1/(tau2*2))
PredCCT[3] <- exp((gamma00+gamma20) + ((gamma01+gamma21)*(2)) +
  (eta*(4)) + 1/(tau2*2))
PredCCT[4] <- exp((gamma00+gamma20) + ((gamma01+gamma21)*(3)) +
  (eta*(9)) + 1/(tau2*2))
PredCCT[5] <- exp((gamma00+gamma20) + ((gamma01+gamma21)*(4)) +
  (eta*(16)) + 1/(tau2*2))
PredCCT[6] <- exp((gamma00+gamma20) + ((gamma01+gamma21)*(5)) +
  (eta*(25)) + 1/(tau2*2))
PredCCT[7] <- exp((gamma00+gamma20) + ((gamma01+gamma21)*(6)) +
  (eta*(36)) + 1/(tau2*2))

INBSTvsCon <- (PredST[2] + PredST[3] + PredST[4] + PredST[5] +
PredST[6] + PredST[7]) - (PredCon[2] + PredCon[3] +
PredCon[4] + PredCon[5] + PredCon[6] + PredCon[7])
INBSTvsCCT <- (PredST[2] + PredST[3] + PredST[4] + PredST[5] +
PredST[6] + PredST[7]) - (PredCCT[2] + PredCCT[3] +
```



```
PredCCT[4] + PredCCT[5] + PredCCT[6] + PredCCT[7])
INBCCTvsCon <- (PredCCT[2] + PredCCT[3] + PredCCT[4] + PredCCT[5] +
PredCCT[6] + PredCCT[7]) - (PredCon[2] + PredCon[3] +
PredCon[4] + PredCon[5] + PredCon[6] + PredCon[7])

INMBConvsST <- (PredCon[2] + PredCon[3] + PredCon[4] + PredCon[5] +
PredCon[6] + PredCon[7]) - (PredST[2] + PredST[3] +
PredST[4] + PredST[5] + PredST[6] + PredST[7])
INMBCCTvsST <- (PredCCT[2] + PredCCT[3] + PredCCT[4] + PredCCT[5] +
PredCCT[6] + PredCCT[7]) - (PredST[2] + PredST[3] +
PredST[4] + PredST[5] + PredST[6] + PredST[7])
INMBConvsCCT <- (PredCon[2] + PredCon[3] + PredCon[4] + PredCon[5] +
PredCon[6] + PredCon[7]) - (PredCCT[2] + PredCCT[3] +
PredCCT[4] + PredCCT[5] + PredCCT[6] + PredCCT[7])

CEACST <- step(-INMBSTvsCon)*step(-INMBSTvsCCT)
CEACCCT <- step(-INMBCCTvsST)*step(-INMBCCTvsCon)
CEACCon <- step(-INMBConvsST)*step(-INMBConvsCCT)

}
```

10.31 QGC Random Intercept and Slope, Fixed Quadratic Term: Predictor for Intercept and Slope, Gamma Distributed Data

```
model{

for (i in 1:N) {

LS[i,1:2]~dmnorm(MU[i,1:2], covmat[1:2,1:2])
MU[i,1] <- gamma00 + gamma10*D1[i] + gamma20*D2[i]
MU[i,2] <- gamma01 + gamma11*D1[i] + gamma21*D2[i]
alpha[i] <-LS[i,1]
beta[i] <- LS[i,2]

for(t in 1:T){
```

```
DV[i,t] ~ dgamma(shape, rate[i,t])
log(mu[i,t]) <- alpha[i] + beta[i]*(Occasion[t]-1) +
   eta*pow((Occasion[t]-1),2)
rate[i,t]<- shape / mu[i,t]

}
}

shape ~ dunif(0, 10000)

gamma00 ~ dnorm(0, .0000000001)
gamma10 ~ dnorm(0, .0000000001)
gamma20 ~ dnorm(0, .0000000001)
gamma01 ~ dnorm(0, .0000000001)
gamma11 ~ dnorm(0, .0000000001)
gamma21 ~ dnorm(0, .0000000001)
eta ~ dnorm(0, .0000000001)

covmat[1:2,1:2] ~ dwish(R[,],2)
R[1,1] <- 3
R[2,2] <- 2
R[1,2] <- 0
R[2,1]<- R[1,2]
Cov[1:2,1:2] <- inverse(covmat[,])
rho1 <- Cov[1,2]/sqrt(Cov[1,1]*Cov[2,2])


PredCon[1] <- exp(gamma00 + (gamma01*(0)) + (eta*(0)))
PredCon[2] <- exp(gamma00 + (gamma01*(1)) + (eta*(1)))
PredCon[3] <- exp(gamma00 + (gamma01*(2)) + (eta*(4)))
PredCon[4] <- exp(gamma00 + (gamma01*(3)) + (eta*(9)))
PredCon[5] <- exp(gamma00 + (gamma01*(4)) + (eta*(16)))
PredCon[6] <- exp(gamma00 + (gamma01*(5)) + (eta*(25)))
PredCon[7] <- exp(gamma00 + (gamma01*(6)) + (eta*(36)))

PredST[1] <- exp((gamma00+gamma10) + ((gamma01+gamma11)*(0)) +
 (eta*(0)))
PredST[2] <- exp((gamma00+gamma10) + ((gamma01+gamma11)*(1)) +
```



```
 (eta*(1)))
PredST[3] <- exp((gamma00+gamma10) + ((gamma01+gamma11)*(2)) +
 (eta*(4)))
PredST[4] <- exp((gamma00+gamma10) + ((gamma01+gamma11)*(3)) +
 (eta*(9)))
PredST[5] <- exp((gamma00+gamma10) + ((gamma01+gamma11)*(4)) +
 (eta*(16)))
PredST[6] <- exp((gamma00+gamma10) + ((gamma01+gamma11)*(5)) +
 (eta*(25)))
PredST[7] <- exp((gamma00+gamma10) + ((gamma01+gamma11)*(6)) +
 (eta*(36)))

PredCCT[1] <- exp((gamma00+gamma20) + ((gamma01+gamma21)*(0)) +
  (eta*(0)))
PredCCT[2] <- exp((gamma00+gamma20) + ((gamma01+gamma21)*(1)) +
  (eta*(1)))
PredCCT[3] <- exp((gamma00+gamma20) + ((gamma01+gamma21)*(2)) +
  (eta*(4)))
PredCCT[4] <- exp((gamma00+gamma20) + ((gamma01+gamma21)*(3)) +
  (eta*(9)))
PredCCT[5] <- exp((gamma00+gamma20) + ((gamma01+gamma21)*(4)) +
  (eta*(16)))
PredCCT[6] <- exp((gamma00+gamma20) + ((gamma01+gamma21)*(5)) +
  (eta*(25)))
PredCCT[7] <- exp((gamma00+gamma20) + ((gamma01+gamma21)*(6)) +
  (eta*(36)))

INBSTvsCon <- (PredST[2] + PredST[3] + PredST[4] + PredST[5] +
PredST[6] + PredST[7]) - (PredCon[2] + PredCon[3] +
PredCon[4] + PredCon[5] + PredCon[6] + PredCon[7])
INBSTvsCCT <- (PredST[2] + PredST[3] + PredST[4] + PredST[5] +
PredST[6] + PredST[7]) - (PredCCT[2] + PredCCT[3] +
PredCCT[4] + PredCCT[5] + PredCCT[6] + PredCCT[7])
INBCCTvsCon <- (PredCCT[2] + PredCCT[3] + PredCCT[4] + PredCCT[5] +
PredCCT[6] + PredCCT[7]) - (PredCon[2] + PredCon[3] +
PredCon[4] + PredCon[5] + PredCon[6] + PredCon[7])

INMBConvsST <- (PredCon[2] + PredCon[3] + PredCon[4] + PredCon[5] +
```

```
PredCon[6] + PredCon[7]) - (PredST[2] + PredST[3] +
PredST[4] + PredST[5] + PredST[6] + PredST[7])
INMBCCTvsST <- (PredCCT[2] + PredCCT[3] + PredCCT[4] + PredCCT[5] +
PredCCT[6] + PredCCT[7]) - (PredST[2] + PredST[3] +
PredST[4] + PredST[5] + PredST[6] + PredST[7])
INMBConvsCCT <- (PredCon[2] + PredCon[3] + PredCon[4] + PredCon[5] +
PredCon[6] + PredCon[7]) - (PredCCT[2] + PredCCT[3] +
PredCCT[4] + PredCCT[5] + PredCCT[6] + PredCCT[7])

CEACST <- step(-1*INMBSTvsTAU)*step(-1*INMBSTvsCCT)
CEACCCT <- step(-1*INMBCCTvsST)*step(-1*INMBCCTvsTAU)
CEACCon <- step(-1*INMBTAUvsST)*step(-1*INMBTAUvsCCT)

}
```

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. Petrov & F. Caski (Eds.), *Proceedings of the second international symposium on information theory*.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2004). *Bayesian data analysis*. New York, NY: Chapman & Hall.

Hamaker, E. L. (2005). Conditions for the equivalence of the autoregressive latent trajectory model and a latent growth curve model with autoregressive disturbances. *Sociological Methods and Research*, *33*, 404-416.

Jongerling, J., & Hamaker, E. L. (2011). On the trajectories of the predetermined ALT model: What are we really modeling? *Structural Equation Modeling: A Multidisciplinary Journal*, *18*, 370-382.

Lunn, D., Jackson, C., Best, N., Thomas, A., & D., S. (2013). *The bugs book: A practical introduction to bayesian analysis*. Chapman and Hall.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). Winbugs – a Bayesian modeling framework: concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325-337.

Plummer, M. (2012). Jags version 3.3.0 user manual [Computer software manual].

R Core Team. (2014). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461-464.

Spiegelhalter, D., Best, N., Carlin, B., & Linde, A. v. d. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *64(4)*, 583-639.

Thompson, S., Nixon, R., & Grieve, R. (2006). Addressing the issues that arise in analysing multicentre cost data, with application to a multinational study. *Journal of Health Economics*, *25*, 1015-1028.

Wetzelaer, P., Jongerling, J., Arntz, A., & Evers, S. (2017). Bayesian multilevel net benefit regression for longitudinal cost-effectiveness data. (submitted)

Yu-Sung, S., & Masanao, Y. (2014). R2jags: A package for running jags from r [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=R2jags` (R package version 0.04-01)

Curriculum Vitae

Curriculum Vitae

Personalia

Name: Pim Wetzelaer
Date and place of birth: April 13th 1982, Kerkrade
E-mail: pwetzelaer@gmail.com

Education & training

2017	Basic Teaching Qualification in Higher Education	Maastricht University
2011	Master in Health Policy, Innovation & Management	Maastricht University
2006	Master in Neurobiology	University of Amsterdam
2006	Master in Cognitive Science	University of Amsterdam
2003	Bachelor in Biomedical Sciences	University of Amsterdam
2000	Gymnasium	Grotius College Heerlen

Research projects

2012-2018	An international, multicentre RCT on group schema therapy for borderline personality disorder	Maastricht University, Faculty of Psychology and Neuroscience
	Trial-based and model-based economic evaluations, development of Bayesian multilevel models for longitudinal cost-effectiveness data and systematic literature reviews (PhD candidate).	

2011	Management of hardware related infections after DBS surgery: A cost analysis Analysis of the average cost savings that can be realized by treating hardware-related infections after DBS surgery first with antibiotics instead of immediate surgical hardware removal (Research internship).	Maastricht University, Faculty of Health, Medicine and Life sciences
2006-2008	Electrophysiology of striatal medium spiny neurons in Parkinsonian rats *In vitro* intrinsic excitability and corticostriatal synaptic plasticity in striatal MSNs in dopamine-depleted rats and transgenic mice using the perforated patch clamp technique and fluorescence microscopy (Research fellowship).	Universite Libre de Bruxelles, Faculty of Medicine
2005-2006	Pilot study on combining *in vivo* multi-unit recording with reverse microdialysis in rat orbitofrontal cortex Assistance in surgical placement and recording sessions, histological procedures and data analysis (Research internship).	Netherlands Institute for Neuroscience, Amsterdam
2004-2005	Electrophysiology and pharmacology of rat dopaminergic neurons *In vitro* electrophysiological studies on the effects of a dopamine D_2 antagonist and a serotonin 5-HT_{1A} agonist on the firing of rat midbrain dopaminergic neurons (Research internship).	Swammerdam Institute for Life Sciences, Amsterdam

Other work experience

2012-2017	Teacher at university	Maastricht University
2010-2011	Administrative and logistics assistant for a commercial distributor of medical equipment	MediServe, Kerkrade
2010	Administrative employee of the tax and customs administration Netherlands	Belastingdienst, Heerlen
2009	Employee of internet pharmacy	DocMorris, Heerlen

Teaching experience

Interim coordinatorship and lecturer:

- Action (2015-2016)

Tutorships in Bachelor 'Psychology' at Maastricht University:

- Body and behaviour (2012-2013; 2013-2014; 2014-2015; 2015-2016; 2016-2017)
- Skills II (2012-2013; 2013-2014; 2014-2015; 2015-2016)
- Perception (2013-2014; 2015-2016)
- Consciousness (2014-2015)
- Action (2012-2013; 2013-2014; 2014-2015; 2015-2016)
- Motivation and emotion (2013-2014; 2014-2015; 2015-2016)
- Evolution and genetics in psychology (2014-2015; 2015-2016)
- Development (2015-2016)
- Methods and techniques (2016-2017)

Tutorships in 'Maastricht Science Program':

- Neuroscience of Action (2014-2015)

Practical assistant in Bachelor 'Psychology' at Maastricht University:

- Neuroanatomy (2016-2017)

Tutorships in Bachelor 'Health Sciences' at Maastricht University:

- Health care in theory (2012-2013)
- Health and interventions (2012-2013)

Tutorship in Bachelor 'GGK' at Maastricht University:

- Stemmingsstoornissen (replacement tutor; 2014-2015)

Tutorships in Master 'Health policy, innovation and management' at Maastricht University:

- Economics of health care (2012-2013; 2013-2014; 2014-2015; 2015-2016; 2016-2017)

Tutorships in Master 'Mental Health':

- Personality disorders (replacement tutor; 2013-2014; 2014-2015)

Supervisor of several student theses on various topics in psychology in the Bachelor programme 'Psychology' (2011-2012;2012-2013; 2013-2014; 2014-2015; 2015-2016; 2016-2017) and the Master programme 'Mental Health: Adult Psychopathology' (2013- 2014)

Student mentor: providing periodic guidance and mentoring to individual students throughout their three-year 'Bachelor in Psychology' curriculum (2012-2013; 2013-2014; 2014-2015; 2015-2016)

List of publications

- P. Wetzelaer, T. Bouwens van der Vlis, M. Tönge, L. Ackermans, P. Kubben, S. Evers, E. Kocabiçak, and Y. Temel. Management of hardware related infections after DBS surgery: A cost analysis. *Turkish Neurosurgery*, 2018.
- P. Wetzelaer, J. Lokkerbol, A. Arntz, T. van Asselt, F. Smit, and S. Evers. Cost-effectiveness and budget impact of specialized psychotherapy for borderline personality disorder: A synthesis of the evidence. *Journal of Mental Health Policy and Economics*, 20(4):177- 190, 2017.
- P. Wetzelaer, J. Lokkerbol, A. Arntz, A. van Asselt, and S. Evers. [Cost-effectiveness of psychotherapy for personality disorders; a systematic review on economic evaluation studies]. *Tijdschrift voor Psychiatrie*, 58(10):717-727, 2016.
- L. L. Bamelis, A. Arntz, P. Wetzelaer, R. Verdoorn, and S. M. Evers. Economic evaluation of schema therapy and clarification-oriented psychotherapy for personality disorders: a multicenter, randomized controlled trial. *Journal of Clinical Psychiatry*, 76(11):e1432-40, 2015.
- P. Wetzelaer, S. Farrell, J.M.and Evers, G. A. Jacob, C. W. Lee, O. Brand, G. van Breukelen, E. Fassbinder, H. Fretwell, R. Harper, A. Lavender, G. Lockwood, I. A. Malogiannis, U. Schweiger, H. Startup, T. Stevenson, G. Zarbock, and A. Arntz. Design of an international multicentre RCT on group schema therapy for borderline personality disorder. *BMC Psychiatry*, 14(1):319, 2014.
- K. Azdad, M. Chàvez, P. Don Bischop, P. Wetzelaer, B. Marescau, P. P. De Deyn, D. Gall, and S. N. Schiffmann. Homeostatic plasticity of striatal neurons intrinsic excitability following dopamine depletion. *PloS one*, 4(9):e6908, 2009.

Presentations

- Eleventh Workshop on Costs and Assessment in Psychiatry, Venice, March 22nd-24th, 2013, Titles of presentations: Issues in the Analysis of Longitudinal Cost-effectiveness Data (I) & Economic Evaluation of Schema Therapy for Personality Disorders (II)
- International Health Economics Association, Milan, Italy, July 12th-15th 2015, Title of presentation: Bayesian net benefit regression on longitudinal cost-effectiveness data from a multicentre RCT
- NVTAG zomersymposium, July 2nd 2015, Utrecht, Title of presentation: Bayesian net benefit regression on longitudinal cost-effectiveness data from a multicentre RCT

- Colloquium Methodology and Statistics, department of Methodology and Statistics, Maastricht University, September 15th 2015, Title of presentation: Bayesian net benefit regression on longitudinal costeffectiveness data from multicentre RCTs

Postgraduate courses

- 'Stereotaxy and Deep Brain Stimulation', March 1st -2nd 2007, Maastricht
- 'Statistics I', February 13th-March 26th 2012, Maastricht
- 'Multilevel analysis of longitudinal data', March 13th-April 12th 2012, Maastricht
- 'Statistics II', April 13th-June 22nd 2012, Maastricht
- EPP statistical course 'Practical data analysis', October 4th-5th 2012, Heeze
- 'WinBUGS in Health Economic Evaluation', October 28th-29th 2013, Cambridge
- EPP advanced statistical course 'Meta analysis', November 7th-8th 2013, Heeze
- 'Bayesian Cognitive Modeling', August 11th -15th 2014, Amsterdam
- EPP workshop 'Programming', February 12th -13th 2015, Heeze
- EPP advanced statistical course 'Multilevel analysis', October 1st -2nd 2015, Heeze
- Workshop 'Discretely integrated condition event (DICE) simulation', February 11th 2017, Maastricht
- EPP workshop 'Network analysis', February 23rd -24th 2017, Heeze

Acknowledgements

Acknowledgements

Many thanks to my supervisors, Prof. Arnoud Arntz, Prof. Silvia Evers, and Prof. Marleen Rijkeboer. I am very grateful for the opportunity to conduct my Ph.D. research under your guidance. It was a true blessing to be guided by your expertise.

Many thanks to my girlfriend Linda for all the love, support, patience, and confidence. Many thanks to my friends and family. Mom and dad, thanks for all the support over the years.

Many thanks to all my colleagues from UM, FPN, EPP, CPS, and HSR. Thanks to the members of the CP section and MTA seminar group for your interesting presentations, nice discussions, and critical comments. Thanks to all the co-authors, trial participants, therapists, site coordinators, student research assistants (Ymke and Dorien, thanks so much!), central research assistants (Michelle, thanks for all your help!), local research assistants, and secretaries. Thanks to Rosanne for data management support.

Many thanks to Dr. Joran Lokkerbol for your guidance in performing a model-based economic evaluation. Many thanks to Dr. Joran Jongerling for your guidance in performing Bayesian multilevel modelling. Many thanks to Prof. Gerard van Breukelen for statistical advice.

Many thanks to Prof. Amanda Kaas for giving me the opportunity to coordinate the 'Action' course, and to Prof. Peter Stiers as well for all the support. Thanks to Noël and Jordi for support with PC issues. Thanks to Lotte Bamelis for the opportunity to discuss, learn from, and build further on the analysis of your data.

Superspecial thanks to Lotte Lemmens and Marjolein, you're two of the kindest people I know. Thanks for the nice lunchbreaks Linda van Zutphen, Lorraine, Martijn, and Ramon. Thanks for the good times and having put up with me as an office mate Glynis, Ken, Thomas, Anna, Elke, Rilana, Irena, Ivan, and Alfons! Thanks to everyone I forgot to mention here.